

# Illuminating the black box

Note to biologists: submissions to *Nature* should contain complete descriptions of materials and reagents used.

**T**his journal aims to publish papers that are not only interesting and thought-provoking, but reproducible and useful. In order to do this, novel materials and reagents need to be carefully described and readily available to interested scientists.

That might seem obvious. But despite the efforts of our editors and referees, papers in the biological sciences are still being submitted — and occasionally published — that do not adequately describe the reagents used. Unless efforts are redoubled to eliminate this practice, we could see an era of ‘black box’ biology, in which outside researchers cannot work out what was done in an experiment.

Some of these ‘rogue reagents’ are the offspring of well-established technologies that have fallen under commercial control. A frequent example involves antibodies, reagents used throughout biology. The standard for publication should be that the authors can show an antibody’s specificity and utility. But this information is sometimes missing from published papers, and readers must rely on a description from a company with a stake in the antibody in question.

But the problem is most acute in the case of new technologies, which sometimes experience a period of rapid development during which the rules just don’t seem to apply. The desire to exploit the latest innovation can outstrip a field’s ability to impose a common standard for description and documentation.

## Strictly business

An example is RNA interference (RNAi), which allows researchers to manipulate gene expression. The critical reagents in RNAi studies are short interfering RNAs, the sequences of which determine which genes are affected. The approach has unleashed a frenzy of papers describing RNAi strategies that illuminate questions in biological systems and could have commercial application.

Manuscripts submitted from the RNAi field show a pattern typical of such new technologies. As the technology becomes commercialized, transparency slips. *Nature* journals have reviewed too many submissions with reagents incompletely described, or where material-transfer agreements limited availability.

Authors face conflicting pressures. Journals set publication standards for openness and availability, but many researchers have personal business interests, universities want to turn research into patent revenue, and biotechnology companies want to keep as much as possible under wraps.

In one case, an author published a paper without sequences of the RNAi reagents used. Editors and referees missed this. A year or so after publication, the author could not meet requests for his reagents, or provide the sequences for his short interfering RNAs. When the journal editors learned of this they asked for the sequences to be added to the paper. But the author said that patent issues and a business deal prevented him from doing so. The author’s lab could not produce the short interfering RNAs, and a company that he had

established didn’t want the author to reveal the sequences, as this would jeopardize its *raison d’être*. This kind of stalemate matters, because it prevents the replication of experiments and inhibits the selection of appropriate controls in subsequent work.

Some authors claim replication is possible without full sequence information or the details of novel compounds. They say that the materials in question are for sale, enabling anyone to duplicate the paper. This misses the point. Scientific progress revolves around producing data that can drive the next stage of investigation. Even if consistent results can be achieved with a black-box reagent, knowing the composition and sequence can be the difference between making an observation and drawing insightful conclusions about it.

## False premise

Another specious claim is that putting the structures of chemical compounds or sequences for short interfering RNAs in a manuscript is unnecessary, because if appropriate controls are described other investigators will know how to control their experiments. This is a false premise: the controls for an experiment designed to test one hypothesis are likely to be different from those needed to tackle another problem. To design appropriate controls one must know

what to control for — information that comes only from structures or sequences.

A troubling situation can also arise when a paper is published and the reagent provided, but a project ending or a person moving ends this provision. Journals have to decide

**“The desire to exploit the latest innovation can outstrip a field’s ability to impose a common standard for description and documentation.”**

whether to push for continued provision, and, if so, how. Is there a sensible statute of limitations on an author’s commitment to provide? Input from the community would help publishers make a sensible policy; comments can be sent to authors@nature.com.

In the meantime, researchers must, when they are arranging the commercialization of their work, bear in mind the implications that these deals may have on their freedom to publish to the standards that the community is entitled to expect.

*Nature* will continue to require that, at acceptance, materials should be fully described in the paper or in the online Supplementary Information. If limited disclosure is noted between acceptance and publication, refusal to provide information may forfeit any offer to publish. Non-disclosure or inaccessibility brought to the editor’s attention after publication will be investigated and actions taken, including, if necessary, public rebukes posted online and appended to the archived paper. Authors are expected to make reasonable arrangements, described accurately in the published work, by which readers can obtain the reagents described in their papers. ■



# Nuclear stalemates

The nuclear powers are maintaining their ageing stockpiles, without much thought or explanation.

Sixteen years after the end of the cold war, roughly 27,000 bombs and warheads are gathering dust in the hinterlands of the world's established nuclear powers. More than 95% of these weapons are in the United States or Russia, but France, China and Britain all have stockpiles of several hundred devices. Each nation spends several billion dollars a year to house and maintain the weapons and train physicists and engineers with weapons expertise. Yet their political leaders have given little thought as to why this approach makes sense, or where it will lead.

During the cold war, the warheads were supposed to deter either a conventional military attack or a nuclear first strike by the opposing side. Under the cheerfully and accurately named paradigm of mutually assured destruction, or MAD, they were made ready for use in large-scale retaliation, to annihilate cities and nations.

That era is, mercifully, behind us. Now the five main nuclear powers all wish to maintain their ageing stockpiles, while offering confused messages about what they are for, and how many are needed.

The United States' latest nuclear-weapons research programme, called the Reliable Replacement Warhead (RRW), captures the flavour of this era of nuclear befuddlement. As described on page 18, the programme's supporters envisage a future nuclear weapon that will be cheap to build and easy to maintain.

They portray the design as neither old nor new, but rather as an amalgam of previous designs. It will be assured to work, but, the designers envisage, will never be tested to see if it does. Calling the weapon 'reliable' angers critics, who hold that existing warheads will

be reliable for many decades. They see the RRW as another jobs programme for the vast US weapons labs, and fear that its arrival will further undermine the Nuclear Non-proliferation Treaty.

The RRW's advocates believe it will allow scientists and engineers to confront some important issues. At the end of the cold war, the United States closed its huge facilities for producing nuclear weapons. At some stage, they say, the country has to decide how to replace its existing weapons.

What's lacking is a political vision to guide such decisions. The administration of George W. Bush has reduced its nuclear-weapons stockpile, as agreed with Russia in 2002. But it has also floated the desirability of nuclear strikes on buried targets. Were it not for its rejection by Congress, the administration would be pursuing new 'bunker-buster' weapons for this purpose. Critics are left wondering whether the RRW is a pathway to

**"Business as usual meets neither military needs nor obligations on disarmament."**

a smaller arsenal, or to a new nuclear weapon.

The United States is not alone in sending mixed messages. France is reconfiguring its arsenal to be 'flexible', while the British government seems intent upon replacing its fleet of Trident submarines. Russia, like the United States, is committed to cutting back its arsenal, although President Vladimir Putin has begun to significantly increase funds for its nuclear-weapons laboratories.

This business-as-usual approach meets neither the nuclear powers' military requirements, nor their obligations, under the 1968 non-proliferation treaty, to move towards nuclear disarmament.

It is past time for these governments to re-examine what their nuclear-weapons research and development capability, as well as their weapons stockpiles, are actually for. Once they've figured that out, it may even be possible for them to manage their arsenals in ways marginally saner than those necessitated by the cold war. ■

# Learning from Africa

Development projects need data as well as money.

The issues that hinder development in sub-Saharan Africa are many and complex, but one factor that stands out for scientists is the dearth of reliable data on the decades of development projects there.

A lack of information on what has worked and what hasn't has contributed to a lack of accountability among donor nations, host nations and even development professionals. Donors in particular have learnt little from past mistakes, and are impatient. When a project fails, as so many do, the tendency has been to move straight on to the next idea.

Development specialists know this, and today data and analysis are prized. In this issue (page 22) we examine the early progress of one notable experiment in Africa. It involves the support of 12 African Millennium Research Villages, which are receiving a package of interventions, at a maximum cost of US\$110 per person per

year, tailored to lift them out of poverty and onto a sustainable path.

The approach has won support from the African governments involved and from private philanthropists, who have pledged \$100 million to a charity, called Millennium Promise, that aims to expand the programme to an additional 78 villages in the next year.

The administrators of the village projects intend to measure 27 important indicators of project performance, mainly by closely monitoring the progress of some 300 households in each village.

They hope to learn three things: whether each intervention works, whether the links between various interventions can be exploited,

**"When a project fails, the tendency has been to move straight on to the next idea."**

and whether the community is ultimately better placed to manage its own future. This last involves 'softer' measures of capacity and sustainability, and will be the hardest both to monitor and to achieve.

It is early days yet — the longest-running project, at Sauri in Kenya, is just two years old — and few hard data are available so far. But it is crucial that the schemes deliver on their research goals and that they absorb lessons, positive or negative, from the data. ■

# RESEARCH HIGHLIGHTS

## Fast evolver

*Proc. Natl Acad. Sci. USA* **103**, 10334–10339 (2006)

When the Andes mountains were pushed up to near their final height between 2 million and 4 million years ago, it created new habitats for hardy plants. Colin Hughes and Ruth Eastwood of Oxford University, UK, chart how lupins, such as those pictured, took advantage of the opportunity.

They show that one species from the genus *Lupinus*, which arrived in the Andes around 1.5 million years ago, has since diversified into 81 different species. This suggests that each new branch of its phylogenetic tree divides, on average, every 320,000 years — a species diversification rate that makes it the fastest-evolving plant group discovered so far. Only the cichlid fish of East Africa's rift lakes are known to be evolving more quickly.



C. HUGHES

## GEOLOGY

### Eruption frozen in time

*J. Geophys. Res.* **111**, D12107 (2006)

The timing and magnitude of a medieval volcanic eruption — one of the largest in recent history — have been pinned down by analysis of ice-core records.

The eruption of the South Pacific volcano, Kuwae, in Vanuatu, was probably a single burst in either late 1452 or early 1453, conclude Chaochao Gao of Rutgers State University in New Jersey and her colleagues. They looked at 33 ice cores from Antarctica and Greenland, which reveal the eruption as a spike in sulphate concentration.

They estimate that the eruption released more sulphate (which cools the climate) than any other event in the past 700 years. The results could serve as a reference to improve the dating of ice-core records.

Kinesin proteins are responsible for transporting materials around in cells, guided along protein rods known as microtubules. Giovanni Cappello of the Curie Institute in Paris, France, Maxime Dahan of the Kastler Brossel Laboratory, Paris, and their co-workers filmed them going about their work in living mammalian cells by attaching semiconductor nanoparticles to the motor proteins. The nanoparticles glowed like light bulbs under a fluorescence microscope, allowing the researchers to track each protein's trajectory (pictured below).

The kinesin molecules typically switched between spells of linear motion, when they were attached to microtubules, and random, diffusive motion when they left the tracks.

## RNA INTERFERENCE

### What a turn-off

*Proc. Natl Acad. Sci. USA* **103**, 10456–10460 (2006)

Neuroscientists have succeeded in turning female mice completely off sex by silencing a single receptor in a specific region of the brain.

Donald Pfaff and Sonoko Ogawa of The Rockefeller University in New York and their colleagues used RNA interference to switch off the oestrogen receptor- $\alpha$  in neurons of the hypothalamic ventromedial nucleus — a

region of the brain previously shown to have a role in sexual behaviour. This caused the female mice to angrily reject the advances of males, instead of assuming a receptive posture.

Although genetic mutations can be induced to knock out the receptor completely, this may affect how the mouse's brain develops. Using RNA interference allowed the researchers to study the receptor in normally developed adult mice.

## NEUROBIOLOGY

### Miraculous recovery

*J. Clin. Invest.* **116**, 2005–2011 (2006)

Brain scans of patients who have been in a minimally conscious state (MCS) for long periods are providing new insights into how nerve cells adapt after a traumatic brain injury.

Researchers led by Henning Voss of the Citigroup Biomedical Imaging Center and Weill Medical College of Cornell University in New York studied the damaged brains of two patients using a technique known as diffusion tensor imaging. One patient had regained the ability to talk after a car accident left him in an MCS for 19 years. The other had remained in an MCS for 6 years.

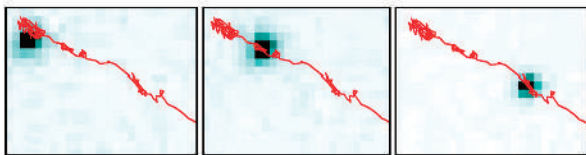
Although large parts of both brains showed damage, the recovered patient's brain revealed regional changes that correlated with clinical improvements. The authors propose that this could represent the regrowth of nerve fibres.

## BIOPHYSICS

### Off the tracks

*Nano Lett.* doi:10.1021/nl060921t (2006)

Motor proteins have been watched as they march along their biomolecular tracks.



AM. CHEM. SOC.



SCIENCE

## ANIMAL BEHAVIOUR

## Mice show empathy

*Science* **312**, 1967–1970 (2006)

A mouse watching a cage-mate writhe in pain will writhe more itself, an observation that Jeffrey Mogil and his team at McGill University in Montreal conclude is evidence of rodent empathy.

The researchers tested mice in twos, giving one or both mildly painful shots of acetic acid. If the two were strangers, they behaved as if they were on their own. But if they had lived together for a few weeks, and both got a shot, they both showed more abdominal constrictions, termed writhing, than when given a shot alone. The effect vanished if the roomies could not see one another.

Empathy has previously been considered an attribute of primates alone. Of course, the empathetic response does not indicate that the mice consciously felt sorry for one another — only that they respond to each other's pain.

## NANOTECHNOLOGY

## Bridging the gap

*Nano Lett.* **6**, 1092–1095 (2006)

Mechanical switches in electrical circuits work by introducing a gap, a behaviour that Marc Bockrath of the California Institute of Technology in Pasadena and his colleagues have replicated at the nanoscale.

The researchers bridged two electrodes with a multiwalled carbon nanotube. Applying a high voltage broke the tube, which put the system into an 'off' state. But the broken ends, the researchers found, could be reunited by charging them. This set up an electrostatic repulsion between the concentric walls of the nanotube, causing the inner tubes to telescope out. The gap was then bridged and the circuit switched on. For



double-walled tubes this switching proved reversible, so it could potentially be used to store information bits in a rewritable memory chip.

## PLANETARY ATMOSPHERES

## Mystery gas

*Geophys. Res. Lett.* **33**, L12301 (2006)

If Earth's volcanoes are a good analogue for those on Mars, researchers will have to look elsewhere for the source of the red planet's mysterious methane.

The detection of methane in the martian atmosphere two years ago led to speculation that it might be produced by some form of life. But non-living sources are possible too — one of the alternatives being martian volcanoes, such as the massive Olympus Mons, a 'shield' volcano that was active until 2 million years ago.

Mauna Loa in Hawaii is Earth's largest shield volcano, which Steven Ryan of the Mauna Loa Observatory and his co-workers have now scanned for evidence of methane emission. In data stretching back to 1987, there is no sign of any methane accompanying the volcanic carbon dioxide.

## PHYSIOLOGY

## Ants take it in their stride

*Science* **312**, 1965–1967 (2006)

Ants, those notorious picnic-crashers, will march long distances in search of food. But how do they find their way home again?

Matthias Wittlinger of the University of Ulm, Germany, and his colleagues show that Saharan desert ants, *Cataglyphis fortis*, use a pedometer to count their strides. The authors allowed a group of ants to march from their nest to an experimental food site. Then, the ants were captured and the researchers either shortened the ants' legs by amputation or elongated them by gluing on stilts made of pig bristles. Both types of altered ants misjudged the distance home — the ants on stumps undershot while the ants on stilts (pictured left) went too far. Further work on the accuracy of the ant pedometer is planned.

## GENETICS

*C. difficile* made easy*Nature Genet.* **38**, 779–786 (2006)

The complete genetic sequence of the troublesome bacterium *Clostridium difficile*, decoded in new work, provides clues to understanding the bug's virulence and emerging drug resistance.

*C. difficile* can cause diarrhoea, which, in the worst cases, leads to death. Researchers led by Julian Parkhill of the Wellcome Trust Sanger Institute in Cambridge, UK, sequenced the 4.3 million base pairs of the circular chromosome of *C. difficile* strain 630, and its smaller plasmid. The chromosome contained an unusually large proportion of mobile elements (11%), which differed in strains extracted from different hosts. They also identified genes associated with antibiotic resistance and virulence factors.

## JOURNAL CLUB

**Morgan Sheng**  
Massachusetts Institute of  
Technology, Cambridge, USA

## A neuroscientist looks at life and death in the brain.

In the mammalian brain, the subject of my research, the same 'use it or lose it' philosophy guides both the life and death of neurons and the strengthening or weakening of the communication junctions between them (the synapses). Recent work

suggests there might be a reason for the parallel.

During brain development, more neurons are born than are needed. Those that compete successfully for limited amounts of growth-sustaining neurotrophic substances, such as brain-derived neurotrophic factor (BDNF), survive.

Synapses are likewise formed in excess during development, and the 'poorly connected' are eliminated, probably also through competition for trophic factors. Active neurons and synapses tend

to win out over inactive ones.

In my lab we study the mechanisms by which some synapses grow bigger and stronger and others are weakened and lost. We already knew that BDNF, in addition to enhancing neuronal survival, promotes the strengthening of synapses.

But Bai Lu and colleagues (N. Woo *et al.* *Nature Neurosci.* **8**, 1069–1077; 2005) have uncovered a new twist. They show that the precursor of BDNF (proBDNF) acts in diametric opposition to BDNF.

BDNF promotes stronger

synapses by acting on the receptor TrkB. Woo and his co-workers find that proBDNF weakens synapses by acting on a distinct receptor, p75NTR. This receptor has previously been implicated in cell death, whereas TrkB has well-known pro-survival activity.

Thus, two different products of the same gene — proBDNF and BDNF — acting on two different receptors elicit opposite outcomes. This leaves me wondering whether neuronal survival and synaptic plasticity are different aspects of similar underlying mechanisms.

## NEWS

# Planet-hunters seek cheap missions

## WASHINGTON DC

Since NASA put its Terrestrial Planet Finder (TPF) missions on indefinite hold earlier this year, researchers have proposed at least four alternative and much cheaper spacecraft that could capture starlight reflected from other worlds beyond our Solar System.

The proposals range from using a giant shield placed 1.5 million kilometres from Earth, as suggested by Webster Cash of the University of Colorado, Boulder, in this week's *Nature* (see page 51), to telescopes with unusual mirrors and finely engineered cameras. The missions have been proposed to NASA's Discovery programme, which caps the costs of its projects at US\$425 million — about a tenth of the cost of NASA's successor to Hubble, the James Webb Space Telescope (JWST).

"The field is more alive than ever," says Alan Boss, a planetary modeller at the Carnegie Institution of Washington.

Astronomers have indirectly detected 194 extrasolar planets so far, all from the ground, by watching the behaviour of their parent stars. But for direct detection, and to stand a chance of studying the planets detected, orbit is the place to be.

The TPF, which would cost much more than the JWST, consisted of two planet-spotting missions that would have picked up small fry the size of Earth. The mission was among the top recommendations of a 2000 National Research Council panel, and featured in President

George W. Bush's vision for space exploration in 2004. But competing priorities caused the agency to decide that the TPF should be "deferred indefinitely" and to delay a smaller planet-hunting Space Interferometry Mission by three years (see *Nature* 439, 768–769; 2006). A similarly ambitious European mission, Darwin, has a proposed launch date of 2015.

But researchers are increasingly optimistic that they can glimpse an alien planet through other methods.

Cash's scheme is to place a specially designed piece of black plastic-bag-like material, 30–50 metres in diameter, between the JWST and a target star, allowing the scope to pick up the faint light of orbiting planets. The shield would have to be about 40,000 kilometres from the telescope — itself around 1.5 million kilometres from Earth — and designed in a distinctive petal shape so that starlight leaking around the edges doesn't drown out the planets. It would have to be able to relocate over great distances to interpose itself between the JWST and various different stars, making it little more than a "big fuel tank with a hefty bag attached", Cash says. But he thinks it could reveal Earth-sized planets around nearby stars.

NASA is now considering Cash's scheme, along with several other ideas, as possible Discovery missions. "In the past five years, there's

been an exponential growth in concepts," says Olivier Guyon, an astronomer at the Subaru telescope on Mauna Kea, Hawaii. Guyon has his own application in with NASA for a 1.2-metre space telescope, which would have a configuration of aspherical mirrors that cut down the light from a star enough to directly observe Jupiter-like planets in nearby systems.

Other teams from NASA laboratories have their own ideas. The Eclipse mission from the Jet Propulsion Laboratory in Pasadena, California, would use an advanced camera to

screen out starlight, while the EPIC mission from the Goddard Space Flight Center in Greenbelt, Maryland, would use interferometry to cancel out starlight, leaving dimmer planets visible. Richard Lyon, EPIC's project scientist at Goddard, says that if his

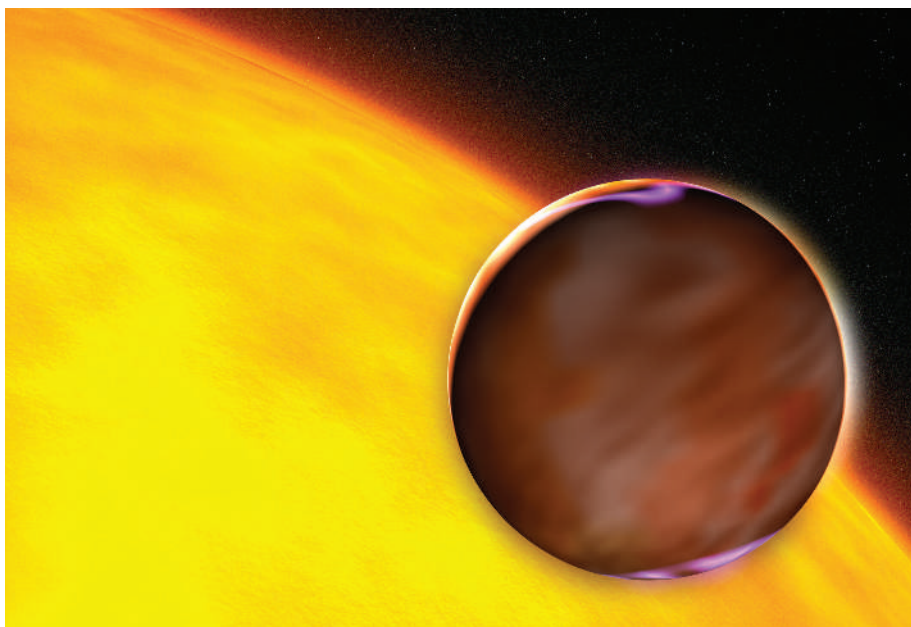
team does not win funding this autumn, they might be able to make the mission even cheaper by shrinking the telescope to fit inside a 1.5-metre-long tube just 45 centimetres in diameter. But a telescope that small, he notes, would have to stare at a single star for almost a year before gathering enough light to detect a planet.

It's not certain that any of these missions will reach space, says Boss. Because of cutbacks to NASA's science budget, researchers must apply for grants for these missions through the Discovery programme, which is normally reserved for missions within our Solar System. Only one such project — the Kepler mission to search for Earth-sized planets eclipsing distant stars — has won funding that way. Kepler, a 1.4-metre telescope, is set to launch in 2008.

Now TPF itself may have been granted a small reprieve, according to Charles Beichman, the mission's project scientist at the Jet Propulsion Laboratory. A congressional budget committee has tentatively earmarked \$10 million for the programme, after the president's budget request earlier this year would have wiped it out completely. If finalized, the money would keep the mission on life support. "But it's clear that the TPF is going to be a long time in the future," says Beichman.

Meanwhile, several ground-based searches are in the planning stages or under way. And in October, the French space agency plans to launch a 30-centimetre space telescope called Corot, with a mission similar to Kepler's. "If we were relying solely on [approved] NASA space missions, the field would have screeched to a halt," says Boss. "But thank God we're not." ■

Geoff Brumfiel



Hope is growing that Earth-like planets beyond our Solar System might be detected by new methods.

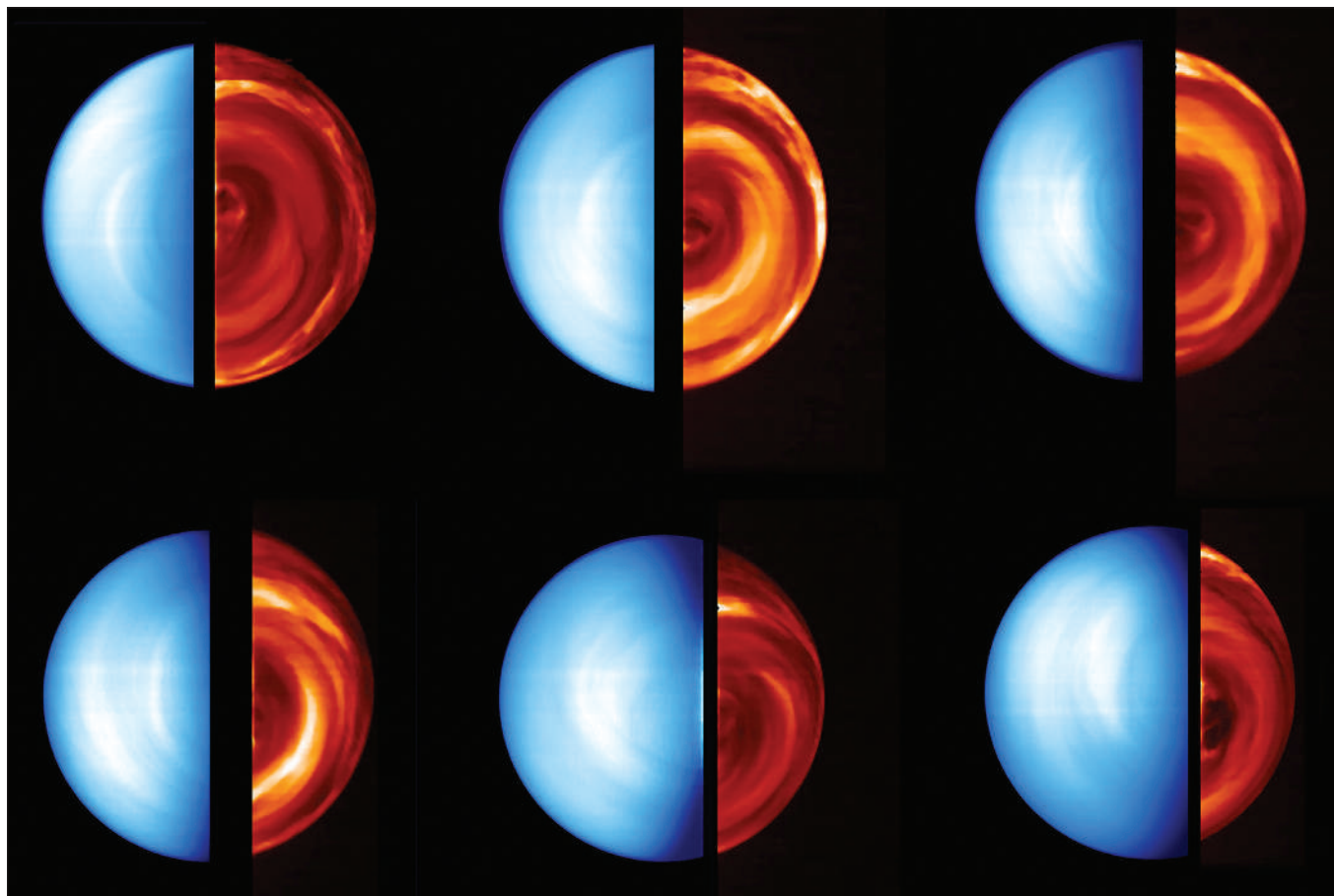
NASA/ESA/G. BACON (GTSCD)



**STILL THE FASTEST COMPUTER ON EARTH**  
IBM's BlueGene/L supercomputer grabs first place again in TOP500 list.  
[www.nature.com/news](http://www.nature.com/news)

TSF/LLNL

ESA/VIRTIS/INAF-IASF/OBS. DE PARIS-LESIA



## SNAPSHOT Venus by day and night

These orbs of blue and red are the planet Venus, as seen by the European spacecraft Venus Express during its cruise into orbit. The composite images are among the first data transmitted from the craft, whose initial results will be discussed at a meeting of the Scientific Assembly of the Committee on Space Research (COSPAR), which convenes on 16 July in Beijing.

The gauzy blue to the left is the planet's dayside, imaged with ultraviolet light; it is bright with reflected sunlight. Scientists aren't sure what causes the white stripes — maybe some kind of aerosol particle is absorbing the ultraviolet light, says Håkan Svedhem, Venus Express project scientist at the European Space Agency (ESA). The

angry red swirls to the right are clouds deep in the venusian atmosphere on the planet's night side. They are seen by infrared light, the only type of radiation that makes it out through the dense atmosphere. The clouds appear as dark shadows in the infrared glow of the planet's hot surface.

Unlike Earth's billowing water clouds, the clouds on Venus are made of sulphuric acid. They float in an atmosphere that is mainly carbon dioxide, and are dragged into a spiralling pattern by fierce, westward winds that tear around the planet at hundreds of metres per second. Here, we view the planet from one end, with the clouds circling the south pole.

The sequence of images was captured by the ultraviolet/visible/

near-infrared spectrometer (VIRTIS) on Venus Express between 12 and 19 April, from distances of between 190,000 and 315,000 kilometres while the spacecraft moved along a long ellipse around the planet. Venus Express, which took just five months to make the journey from Earth to its neighbour, has since moved into a closer, final orbit.

The science instruments were switched off during the craft's orbital manoeuvres, but since 4 June, data from its cameras, spectrometers and analysers have been streaming back to ESA's deep-space antenna at Cebreros, just outside Madrid, Spain. Earlier missions to Venus, by NASA's Mariner probes and the Soviet Union's Venera programme among others, have revealed atmospheric

features like those seen in this image. But now the craft is in its final orbit, its suite of modern instruments is probing different aspects of the atmosphere's composition and dynamics in unprecedented detail.

Svedhem says the mission scientists are busy planning the craft's future observations, so they have had little chance to look at the data flooding in. They have two more weeks to do so before presenting their findings so far — on topics ranging from the hunt for surface volcanism to movies of the cloud tops — at a dedicated session in Beijing. "The COSPAR meeting will be the first time we've speculated about what we are seeing," says Svedhem. Jenny Hogan



# Sociologist fools physics judges

After more than 30 years of studying the physicists who work on gravity waves, spending countless hours talking to physicists and writing a book on the history and sociology of the field, social scientist Harry Collins had a question. Could he pass as a physicist?

He reckons he can — and he has the experimental data to prove it. Collins's study, to be published later this year, is the first experiment on the concept of 'interactional expertise', an idea that could influence areas such as peer review and science journalism. It could even help settle a question lingering since the science wars of the 1990s, when sociologists launched what scientists saw as attacks on the very nature of science, and scientists responded in kind.

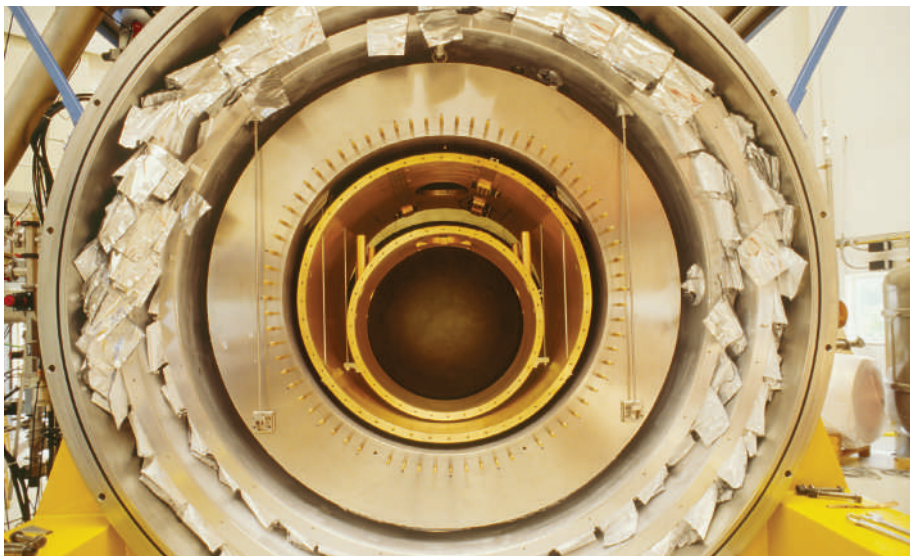
Collins's claim rests on his answers to a set of seven questions about gravity-wave physics set by a gravity-wave physicist. His replies, together with those from a real gravitational physicist, were sent to nine researchers in the field (see 'Faking it'). Asked to spot the real physicist, seven were unsure and two chose Collins. The results appear in a paper co-authored with Rob Evans, like Collins based at Cardiff University, UK. It is due to be published this December in *Studies in the History of Philosophy of Science* (for a preprint see [www.cf.ac.uk/socsci/expertise](http://www.cf.ac.uk/socsci/expertise)).

Nature sent the questions and answers to Sheila Rowan, a gravitational-wave physicist at the University of Glasgow. She was likewise unable to spot the impostor. "The answers are different but it's not obvious which are not by a graduate scientist," she says.

"I could not run LIGO [a US gravity-wave detector] or do lots of other things," says Collins. "But the results do show that outsiders can develop a kind of expertise in a scientific field", even if they cannot carry out the relevant experiments and do not know the mathematics involved.

Collins says this kind of expertise, known as interactional to contrast with the 'contributory expertise' that comes from being able to do experiments and develop theories, should not be dismissed. He points out, for example, that it is important in activities such as grant allocation, in which peer-review panels may include scientists who know the concepts associated with a field, but lack technical understanding.

In a second experiment, Collins and Evans got groups of colour-blind people to pretend they could see colours. Judges compared their performance in conversation with that of people with normal sight. As Collins expected, the colour-blind, immersed in the language of



T. GUICCIARDINI/SPL

He might not be able to run this gravity-wave detector, but a social scientist passed himself off as a physicist.

colour vision, had the interactional expertise to pass as colour-perceivers. In contrast, people lacking perfect musical pitch could not pass for those who could. Very few have such an ability, so those lacking it have not learnt to describe how the skill feels.

## Faking it

One answer to the following question is from an experienced gravity-wave physicist, the other is from social scientist Harry Collins of Cardiff University, UK. To find out which is which, see page 15.

**A theorist tells you that she has come up with a theory in which a circular ring of particles is displaced by gravitational waves so that the circular shape remains the same but the size oscillates about a mean size. Would it be possible to measure this effect using a laser interferometer?**

**A:** Yes, but you should analyse the sum of the strains in the two arms, rather than the difference. In fact, you don't even need two arms of an interferometer to detect gravitational waves, provided you can measure the round-trip light travel time along a single arm accurately enough to detect small changes in its length.

**B:** It depends on the direction of the source. There will be no detectable signal if the source lies anywhere on the plane that passes through the centre station and bisects the angle of the two arms. Otherwise there will be a signal, maximized when the source lies along one or other of the two arms.

If the concept of interactional expertise catches on, it could affect the argument about whether an outsider, such as an anthropologist, can properly understand another group, such as a remote rural community. The debate was part of the science wars, when some scientists claimed that sociologists studying science did not understand the disciplines involved, in part because they did not practise them.

Collins's results do not end that discussion, but they do suggest that outsiders can develop expertise in a field. Collins says that investigators now have a way to display their expertise — and that they and their critics can talk sensibly about whether it is appropriate.

One of the main protagonists in the debate was Alan Sokal, a physicist at New York University who authored a spoof science-studies paper that was accepted by *Social Text*, a cultural research journal. The paper, which consisted of meaningless arguments about quantum theory, was intended to expose what Sokal and others saw as a lack of academic rigour among sociologists.

Sokal says he is struck by Collins's skills in physics, but notes that such understanding would not be enough for more ambitious sociology research that attempts to probe how cultural and scientific factors shape science. "If that's your goal you need a knowledge of the field that is virtually, if not fully, at the level of researchers in the field," says Sokal. "Unless you understand the science you can't get into the theories."

Jim Giles



# TOP FIVE SCIENCE BLOGS

Weblogs written by scientists are relatively rare, but some of them are proving popular. Out of 46.7 million blogs indexed by the Technorati blog search engine, five scientists' sites make it into the top 3,500.

**Declan Butler** asks the winners about the reasons for their success.

179th

**Pharyngula**



♦ <http://scienceblogs.com/pharyngula>

Phil Myers, a biologist at the University of Minnesota, Morris, puts his top rank down to "tapping into the broader areas of liberal politics and atheism" and a rich vein of "resentment against the reactionary religious nature of American culture". Scientists can easily translate their expertise into blog posts, adds Myers. "Sometimes, I just summarize some basic concepts as I would in the classroom." But you are certain to fail if you write as if for a peer-reviewed journal. "It doesn't work on the web," says Myers. "A blog's more like the conversation you'd have at the bar after a scientific meeting."

as possible, and a hell of a lot of patience to deal with all those comments rolling in". Gavin Schmidt, at NASA's Goddard Institute for Space Studies in New York, says the blog fills "a hunger for raw but accessible information" that goes deeper than newspaper articles, but is more easily understood than the scientific literature. "Magazines fill a void, but they can't react or interact as effectively as blogs."

2,174th



♦ <http://cosmicvariance.com>

Frequent posting of original content is crucial to building an audience, says Sean Carroll at Cosmic Variance, which is produced by five physicists. But taking "stances that are provocative and make people think" also helps. One needs to become the place to go for a subject, he says. Citing other blogs is a sure-fire way to get their notice and maybe a citation in return, he adds. But he cautions that citation counts and rankings can be a distraction. "It would be a shame if people worried about traffic and not about having a good blog."

1,647th



♦ [www.pandasthumb.org](http://www.pandasthumb.org)

Being a group blog is key, says contributor Jack Krebs, president of Kansas Citizens for Science. "We have some of the most well-informed observers and critics of the 'intelligent design' and creationist movements." The nature of the topic helps too, he adds. "There is an interest, a hunger even, for thoughtful analysis of the issues related to evolution and creationism."

3,429th



♦ <http://scienceblogs.com/scientificactivist>

Nick Anthis, who only began blogging in January, knows the reason for his site's swift rise to fame. During a political censorship row at NASA in February, Anthis was the first to reveal that a key official had lied about graduating from Texas A&M University. "Before I knew it, it had exploded into a major national news story and he resigned." After an initial spike in traffic, many stayed on as regular readers.

1,884th



♦ [www.realclimate.org](http://www.realclimate.org)

Stefan Rahmstorf, a climate scientist who blogs at RealClimate, puts its success down to the hot topic and expert contributors. It helps to have "a passion for explaining things as clearly

## How the blogs were ranked

There's little agreement about how to rank blogs, so this exercise is best viewed as a rough snapshot. *Nature* trawled the web to identify as many science blogs as possible. Although there are many popular blogs produced by science writers, we included only those written by working scientists covering scientific issues. We asked Technorati to give each one a rank — this popular search engine ranks blogs by measuring the number of sites linking to them in the past six months. For more about the ranking or to comment on the results

♦ [http://blogs.nature.com/news/blog/2006/07/top\\_five\\_science\\_blogs.html](http://blogs.nature.com/news/blog/2006/07/top_five_science_blogs.html)

# Simple recipe gives adult cells embryonic powers

## TORONTO

Reprogramming adult human cells to repair damaged tissues or organs may not be quite as tough as thought. Researchers have devised a chemical cocktail that makes adult mouse cells behave like embryonic stem cells, and the recipe is surprisingly simple.

Human embryonic stem cells can develop into all cell types in the body, an ability known as pluripotency, and scientists think these cells will be invaluable for both research and medicine. But the cells are extracted from human embryos, which is controversial in many countries. To skirt the ethical problems, scientists have been searching for alternative sources of such cells (see *Nature* **441**, 1038; 2006). And little excites the field more than the possibility of a chemical recipe that can reprogramme adult cells into an embryonic state — without involving an embryo.

Now, Shinya Yamanaka and his colleagues at Kyoto University, Japan, have developed what looks like a good first draft for this recipe, at least in mice. They say it takes only four chemicals. For their experiment, the researchers chose fibroblasts, adult cells that divide quickly and can already give rise to some other types of cell. When they added the four factors to fibroblasts, the team says the cells looked and behaved a lot like mouse embryonic stem cells. “Potentially, if we use these four factors with human cells, we could avoid the ethical issues and make pluripotent cells,” says Yamanaka.

The reprogrammed cells pass many of the crucial tests of ‘stem-ness’. They express some of the key genes that embryonic stem cells do, can be coaxed to make the three main cell types of the body, and can divide to give more cells like themselves. Yamanaka calls them “embryonic-stem-cell-like cells”.

## Luck and skill

Yamanaka unveiled his research on 30 June at the International Society for Stem Cell Research in Toronto, Canada, revealing that success involved patience, ingenuity and luck. Over five years, his researchers compiled a list of 24 candidate factors that help stem cells stay flexible. They engineered adult mouse cells so that they would be killed by a drug unless they turned on a gene active only in stem cells. Then the team added genes for the 24 candi-

date factors to the engineered mouse cells, and dosed them with the drug. Only the stem-cell-like cells survived.

The researchers repeated this experiment, removing one or a few genes at a time, until they arrived at four essential chemicals. Three of the factors — Oct4, Sox2 and c-Myc — were already known to be important for stem-ness. But the fourth was a surprise, says Yamanaka. Stem-cell research is so competitive that he

**“Yamanaka seems to have hit a home run.”**

refuses to name this fourth factor until he can publish his work in a scientific journal.

His research adds to work on which factors are key to reprogramming. Ihor Lemischka of Princeton University, New Jersey, and his colleagues have studied 70 genes in mouse embryonic stem cells (J. Silva *et al.* *Nature* doi:10.1038/nature04914; 2006). And Austin Smith’s team at the University of Edinburgh, UK, is investigating a protein called Nanog (N. Ivanova *et al.* *Nature* doi:10.1038/nature04915; 2006). “Several researchers had shown factors that were necessary for programming, but nobody had shown which factors were sufficient,” says Yamanaka.

Researchers are impressed and surprised at Yamanaka’s achievement, mainly because he gambled everything on the key factors being included within his pool of 24 candidates. “He seems to have hit a home run,” says Lemischka.

But scientists caution that Yamanaka’s report has not eliminated the need for work on embryonic stem cells. Researchers must test the same four factors in human cells. And it is not entirely clear whether the reprogrammed cells can do everything that embryonic cells can. Although many of the genes they express are the same, many are not.

Yamanaka’s report came just a day after the US Senate said it would vote on relaxing rules on embryonic research later this year. Some have argued that progress in reprogramming has made work on embryonic stem cells unnecessary, and they may seize on Yamanaka’s work to bolster this position. But scientists at the Toronto meeting said that would be a mistake.

“There will be people who say that, and they will be wrong,” says Lemischka. “There’s a lot more work to do to understand these cells. The science is really solid, but it is by no means true that reprogramming has now been solved.” ■

Erika Check

## ON THE RECORD

**“That excommunication goes for the woman, the doctors and the scientists who eliminate the embryo.”**

Cardinal Alfonso López Trujillo, head of the Pontifical Council for the Family, explains how he’d like to see all those involved in embryonic stem-cell research excommunicated from the Catholic Church.

**“I prayed to find a mummy, but when I saw this, I said it’s better, it’s really beautiful.”**

Nadia Lokma, chief curator of the Egyptian Museum in Cairo, says she couldn’t be happier about finding embalming materials and flowers in a coffin opened in Egypt’s Valley of the Kings on 28 June.

Sources: Catholic Online, Washington Post

## SCORECARD



### Giant beetles

Furniture makers in Wales are shocked by a huge bug in their timber. It turns out to be a rare Capricorn beetle, not seen in the country for centuries.



### Stem-cell therapy

StemCells of Palo Alto announces success in treating a mouse model of Batten’s disease, in which neurons lack the enzymes to break down a toxic chemical. Treatment with human fetal stem cells prolongs the lifespan of mice with this rare and fatal illness.

## NUMBER CRUNCH

Good news for Britain’s biomedical researchers: the finances of the Wellcome Trust are looking very healthy indeed. The trust, a medical charity that has funding clout on a par with the UK government, said on 3 July that it will raise even more money by issuing its own bonds.

**£12.3 billion** is the amount in the charity’s endowment, up from £8.1 billion (US\$14.9 billion) in 1996.

**10%** is the average annual return earned by the trust’s fund managers over the past ten years.

**£500 million** in one-off payments is the sum the charity hopes to earn by issuing bonds.



R. AUSTING/FLPA

# Doing conservation by numbers

## NEW YORK

Tiger conservationists are trying a more business-like approach to saving the threatened species — focusing not on collecting money but on performance standards.

Biologists at the Wildlife Conservation Society (WCS) in New York want to increase the number of tigers (*Panthera tigris*) at the society's research sites in Asia by 50% over the next decade. WCS director for science Alan Rabinowitz says the plan, called 'Tigers Forever', will differ from previous conservation attempts because the society's biologists, stationed across Asia, have set targets by which their success can be judged.

In the past, conservation groups have issued pleas for funds to save tigers while often having no clear idea of how that aim should be achieved. It is virtually unprecedented for a group to announce performance targets in advance. "This is not a fund-raising exercise," says Rabinowitz. "We're putting our ass on the line here. Tigers Forever is designed to measure our impact on increasing tiger populations at specific sites."

As well as judging overall success, the information gained about what works and what doesn't in different areas will guide future action. Although conservation groups have made strides in boosting tiger numbers in some areas,

particularly India and Russia, "we are not using the lessons learned", says Rabinowitz. Ullas Karanth, an Indian biologist who works with the WCS, adds that "we are trying to be tough-minded with our people, seeking rigorous reviews of tiger plans".

That philosophy seems to be attracting donors. Michael Cline, a New York businessman and WCS board member, has contributed to the US\$10 million already pledged to Tigers Forever. "More organizations should set goals like this," he says. "As a venture capitalist, I believe in fact-based judgements based on likelihood of success."

"I definitely think this will become a trend," adds Sybille Klenzendorf, a biologist who coordinates tiger programmes at the World Wildlife Fund in Washington DC.

From the forests of the Asian subcontinent to the steppes of Russia, poaching, illegal hunting and habitat loss have reduced tiger numbers

dramatically in past decades, to between 3,000 and 5,000 animals (see *Nature* **441**, 927–930; 2006). The situation is most critical in the jungles of southeast Asian nations, and the WCS will begin by concentrating on selected reserves there (see 'Wildlife Conservation Society Sites in Asia').

One of the principal goals of the programme is to increase tiger prey. "Tigers eat large, hoofed animals," says Karanth. "The key to saving tigers is to maintain proper prey density."

WCS biologists in Mondulkiri are relatively optimistic despite the low tiger numbers. Prey species have been dramatically reduced by poaching, owing to the many guns available after years of war. But good habitat remains for both prey and tiger. To bring back the prey species, the WCS is encouraging local agencies and governments to set up a variety of incentive-based actions to cut poaching. These include paying informers to expose illegal hunting, offering bounties for guns or tiger traps, and instituting bonuses for diligent rangers.

"These are not nice things that people want to hear about," says Rabinowitz. But he says that measures such as the number of traps turned in or the number of poachers identified "are really good surrogates to show how much better things are".

Rex Dalton

WILDLIFE CONSERVATION SOCIETY SITES IN ASIA

| Country      | Tiger site                           | Area of site (km <sup>2</sup> ) | Estimated number of tigers in site |
|--------------|--------------------------------------|---------------------------------|------------------------------------|
| Cambodia     | Seima Biodiversity Conservation Area | 3,000                           | < 10                               |
| India        | Malend-Mysore Project Landscape      | 5,560                           | 259                                |
| Indonesia    | Gunung Leuser National Park          | 9,500                           | 90                                 |
| Lao PDR      | Nam Et Phou Louey Protected Area     | 4,300                           | 23                                 |
| Myanmar      | Hukaung Tiger Reserve                | 21,800                          | 100                                |
| Thailand     | Huai Kha Khaeng and Thung Yai        | 6,000                           | 120                                |
| <b>Total</b> |                                      | <b>50,160</b>                   | <b>602</b>                         |

The total area of WCS tiger sites in Asia (including Russia and China) is 261,355 km<sup>2</sup>.



**STEM-CELL PODCAST**

Listen to debate and interviews on this hot topic online.

[www.nature.com/podcast/stemcells](http://www.nature.com/podcast/stemcells)

# Should conservation biologists push policies?

## SAN JOSE

For conservation biologists, it's the question that won't go away. Should they make the leap from describing the facts of a case, to telling people what ought to be done?

Biologist Reed Noss of the University of Central Florida believes they should. Addressing a meeting of the Society for Conservation Biology (SCB) in San Jose, California, last week, he tried to convince the crowd that they have a responsibility to be not just scientists dealing in objective facts, but also advocates pushing particular policies.

But Mike Scott of the University of Idaho, organizer of the symposium, thinks the SCB should stick to the facts. "We need to position ourselves as the go-to authority on

conservation matters worldwide," he says. "We can more forcefully do that if we do rigorous science, and then leave it for the decision-makers to figure out what to do with that."

The advocacy question is perhaps more difficult for conservation biologists than many other scientists. Their field is already premised on the value of having lots of species around. And most of these scientists got into the field because of their strong feelings about nature, the wilderness and often particular species.

Peter Brussard, a population biologist at the University of Nevada, Reno, points out that the debate goes back to at least 1951. Then the Nature Conservancy split from the Ecological Society of America because of a dispute over whether

scientists should do more than just describe.

These days, he thinks, "the debate has been reframed a little bit", with more researchers willing to be advocates. In the end, he says, much depends on the definition of advocacy. "We never seem to get beyond semantics."

For example, does advocacy include sending a paper to policy-makers? Or to the press? Or reiterating your findings if you don't think policy-makers have taken enough notice of them?

Back in San Jose, the US Geological Survey's Susan Haseltine warns the meeting of the harm a scientist can do to their credibility by being an activist. "I don't believe you can be strong in science and in advocacy," she says.

Meanwhile, the policy-makers to whom all this advocacy is directed have their own views about the information coming from scientists. "I don't mind policy recommendations from scientists," says Andrew Wheeler, staff director and chief council of the US Senate Committee on Environment and Public Works, which handles conservation legislation. "But I take them with a grain of salt."

Despite the arguments against, it looks as if the members of the SCB are coming round to Noss's point of view. In a survey of about 300 attendees, 70% felt that the voice of the society — the journal *Conservation Biology* — should advocate certain policies. ■

Emma Marris

## Senator backs halfway houses for nuclear waste

A powerful US senator has proposed building temporary storage facilities for nuclear waste around the country.

Frustrated by delays to building a permanent waste repository at Yucca Mountain in Nevada (see *Nature* 440, 987–989; 2006), Senator Pete Domenici (Republican, New Mexico) is advocating temporary “consolidation and preparation” facilities. The proposal, which Domenici included in a Senate appropriations bill last week, would instruct the Department of Energy to draw up plans for the facilities in states with nuclear power stations. Nuclear waste is currently stored on site at the plants.

Germany and Sweden already have such facilities. “There are no technical barriers to interim storage,” says Kevin Crowley, who directs the Nuclear and Radiation Studies Board of the National Academy of Sciences in Washington DC. But, he warns, building and licensing facilities around the country will be “very expensive”.

## Flying telescope takes one step forward, two back

The Stratospheric Observatory for Infrared Astronomy (SOFIA) has got the go-ahead from NASA — but the project is still firmly on the ground.

NASA gave the long-delayed programme, which would deploy a plane carrying a 2.5-metre infrared telescope, the thumbs-up during a crucial review on 15 June. But programme scientists say SOFIA may be tested at NASA's Dryden facility in southern California rather than at Ames Research Center in Moffett Field, California, where the project is based — a change that could push the first science flights back to 2010.

Dana Backman, outreach coordinator for the project, says the team still hopes to be doing science by July 2008. According to NASA, the plane was intended to be operational by 2004. But any delay may become moot if no funds can be found for the project at NASA, which has suffered cuts to its science budget. SOFIA is a collaboration between the agency and the German Aerospace Center.

## Tsunami warning system goes online in Indian Ocean

The Indian Ocean's early-warning system for tsunamis is up and running. The US\$2-million system, set up and coordinated by the United Nations Economic, Scientific and Cultural Organization (UNESCO) after the 2004 tsunami, consists of a seismic network,



Despite an early warning system, people may still be at risk from tsunamis in the Indian Ocean.

a set of buoys, and several anchored deep-sea instruments that measure the power and propagation of waves.

All data are sent in real time to the Pacific tsunami warning centres in Japan and Hawaii. If there is a quake big enough to cause a tsunami in the Indian Ocean, these centres will warn authorities in 24 affected countries. Four nations bordering the Indian Ocean — Somalia, Yemen, Saudi Arabia and the United Arab Emirates — are not yet linked in to the system.

But because the flow of information from local authorities to people at risk is still unresolved, the system may not yet be able to save lives. If a tsunami were to strike tonight, millions living near beaches would not be alerted in time, says UNESCO spokeswoman Sue Williams. An intergovernmental group will discuss this and other problems at a meeting in Bali from 31 July to 2 August.

## US biologist who faked data faces a year in jail

For the first time, a US researcher has received a jail sentence for faking data on grant applications to the National Institutes of Health.

On 28 June, in a federal court in Vermont, physiologist Eric Poehlman was sentenced to

one year and a day after he pleaded guilty last year to making false statements to secure a \$542,000 grant on hormone replacement for women (see *Nature* 434, 424; 2005). He is to report to federal prison by the end of August, unless he appeals against the sentence.

After being accused of fabricating data in 17 grant applications worth \$11.6 million and publishing ten articles with falsified data, Poehlman struck a plea deal for the offences, committed while he was working at the University of Vermont in Burlington and the University of Maryland in Baltimore.

Poehlman, who resigned from his job at the University of Montreal, Canada, early last year, has retracted the articles, returned \$180,000 to the government, and paid \$16,000 in attorney's fees for Walter DeNino, the young researcher who blew the whistle on him.

## Hubble camera working again after 11-day hiatus

One of the Hubble Space Telescope's cameras is scanning the cosmos again, after an electronics glitch shut it down for nearly two weeks.

On 30 June, engineers switched to back-up power for the Advanced Camera for Surveys, which stopped working on 19 June. Hubble's three main other instruments had continued working normally.

No crucial observations were lost while the camera was offline, says David Leckrone of the NASA Goddard Space Flight Center in Greenbelt, Maryland, Hubble's senior project scientist. When it shut down, it had been slated to make follow-up observations on the Hubble Ultra Deep Field, a survey of distant galaxies. That work will now be rescheduled for the autumn.

### Answer from 'Faking it' on page 8

The response from Harry Collins was answer B.

## Villagers and museums wrangle over bear's body

The corpse of Bruno the bear is in high demand, but will probably end up as a teaching tool.

Bruno (right) was part of a programme to reintroduce the brown bear (*Ursus arctos*) to the Italian Alps, but he went rogue and headed north to Bavaria. On 26 June he was shot near an Alpine lake by a hunter.

Gunther von Hagens, creator of exhibitions that display the flayed and plastinated bodies of humans and animals, offered to donate €10,000 (US\$12,800) to an animal-welfare organization for Bruno's flesh. Villages close to where he was shot have also put in claims.

But Bavaria's environment ministry says Bruno will be put to more dignified use, for



“teaching and biological research”. His stuffed fur will be displayed at the Museum for People and Nature in Munich, and his skeleton and inner organs retained for student instruction.

A tissue sample has also been sent for DNA analysis — just to make sure it is the right bear.

B. MARQUEZ/AP

A. HÖTZELSPERGER/DPA

## BUSINESS

# Big Blue in deep

India is traditionally suspicious of multinationals. So IBM's huge investment plan has provoked a mixed response, reports K. S. Jayaraman.

**T**he news that one of the world's largest computer companies is planning to invest a staggering US\$6 billion in India over the next three years has sent ripples of excitement through the nation's emerging information-technology industry.

IBM chairman Samuel Palmisano was joined by the president of India, A.P.J. Abdul Kalam, when he announced the expansion to a cheering crowd of thousands of contractors and employees in Bangalore on 6 June (see *Nature* **441**, 811; 2006).

A recruitment spree over the past three years has already brought the number of IBM's employees in India up to 43,000 — by far its largest contingent outside the United States. So last month's move has got analysts asking just how far the computer company might go in transferring skilled, high-value jobs — including research positions — from Europe and the United States to the subcontinent.

So far, most of IBM's activity in India has involved hiring computer programmers and customer-support specialists, many of whom work at a huge centre in Bangalore. From there they service the needs of client companies who use IBM computers and services around the world.

The company hasn't yet taken the next step of doing research and major product development in India: its activities in this sphere have been relatively modest. That comforts political leaders in the United States and Europe, who fret about such jobs moving abroad. But it frustrates some observers in India, who accuse the computer firm of taking advantage of cheap, skilled labour without relocating high-value positions.

"Job booms created by IBM and similar companies are only a short-term gain for India," says C. N. R. Rao, science adviser to the prime minister, Manmohan Singh. "Instead of making our skills cheaply available to others, we should try to become leaders ourselves."

IBM currently has two research laboratories in India — a communications lab established eight years ago in New Delhi and a smaller one that opened last August in Bangalore, to back up the support centre there. The labs employ



The chairman of IBM is joined by India's president as he reveals major investment in Bangalore.

about 110 people between them. That's about 3% of IBM's research staff, most of whom work at its major corporate labs in Zurich, New York's Yorktown, and Almaden in California.

**"The centre of gravity of global software development is going to shift to India."**

IBM hopes to expand the proportion of its research that supports computer services, says Mahmoud Naghshineh, IBM's head of research in that area. He says that India will play a big part in the expansion. "I hope the Indian labs will grow to be the size we have elsewhere," he says. "But the goal is not to cut down in the United States. We need to grow faster in India to have a more solid impact there."

According to Naghshineh, the great attraction of India is its workforce. "The hard part is that a lot of similar things are happening in other companies," he says, and so staff turnover is very high. "You have to be street-smart in managing your business."

The Bangalore facility is a "living lab" that benefits from direct proximity to the people providing computer services, Naghshineh says. "It will do a lot of applied work. But we have a portfolio to explore new ideas, as well."

The expansion was welcomed by some rivals. "It is good for the community as a whole," says Mathukumalli Vidyasagar, a vice-president of Hyderabad-based Tata Consultancy Services, one of Asia's largest computer-services companies and a major IBM competitor. "If IBM starts doubling its headcount, the centre of gravity of global software development is going to shift to India, giving legitimacy to our industry," Vidyasagar says.

And India's Congress-led government strongly supports the arrival of multinational companies. It sees a promising model for the future in General Electric's John F. Welch Technology Centre in Bangalore — home to 2,200 scientists and engineers, more than half of them with PhDs. But not everyone is so enthusiastic.

IBM left India entirely in 1978, during the long cold-war period when the country viewed investments by multinationals with distrust. It crept back into India in 1992, after economic liberalization, but the company's activities there really took off in 2003, when it made a strategic decision to base support services for global clients in India, and started recruitment there on a large scale.

Now smaller firms are getting nervous about competing with IBM for staff. And research managers fear that the lure of high wages from multinational companies will pluck scientists from their low-paid positions in government laboratories, with negative long-term effects. "We already see this in the space, atomic-energy and defence sectors," says Gangan Pratap, director of the government-run Centre for Mathematical Modelling and Computer Simulation in Bangalore.

IBM has been cagey about the details of its investment plan, and speculation has arisen that much of it will go on acquisitions. Two years ago, the company bought Daksh eServices, a New Delhi back-office outsourcing firm, for \$160 million — and has since increased its workforce from 6,000 to 20,000. According to industry insiders, such purchases are a better route to accruing experienced and reliable staff than attempting to hire them from scratch.

India's computer industry will create 2.3 million jobs by 2010, indirectly spawning a further 6.5 million positions, according to the US consultancy Kelly Services. IBM has clearly decided to be part of this explosion: "India is going to be key for us," says Naghshineh. ■

G. SINGH/AP



# THE NEXT NUKE

US nuclear weapons scientists are designing a warhead that is meant to be 'reliable' without ever having been tested. **Geoff Brumfiel** asks whether it could renew the United States' ageing stockpile.



Wasteland: America's nuclear-weapons test ranges have lain silent since the country declared a testing moratorium in 1992.

NNSA/NEVADA SITE OFFICE

It's a hot spring day in the Nevada desert, and retired technician Ernie Williams is showing tourists how nuclear bombs used to be tested. Williams is short and gruff and a veteran of the US weapons complex. Since 1951, he has participated in around 80 nuclear detonations or 'shots'. By his estimate, he has received 0.6 sieverts of radiation over his lifetime — about 40 times the dose of a modern nuclear-plant worker. But his only ailment is slight deafness, for which he blames genes, not nuclear blasts.

From the tour bus, Williams points to a 15-storey tower rising above the pockmarked landscape of the Nevada test site. The tower was built for 'Icicap', one of the country's last three underground tests, which were cancelled in 1992 after the United States declared a testing moratorium. Workers simply left the tower baking in the desert sun, a monument to decades of nuclear explosions.

Five hundred kilometres west, at California's Lawrence Livermore National Laboratory, Doug East is explaining to visitors the new face of nuclear testing. East is a computer scientist who spent his early career programming large networks at places such as IBM and the telephone giant Pacific Bell. He came to the Livermore laboratory in 1992 — the same year the

United States stopped testing nuclear bombs.

East leads the guests inside a cool, sterile vault that houses the lab's newest supercomputer, known as Purple. An array of black boxes, each stamped with 'IBM', Purple has the electricity needs of a small town and is one of the fastest computers on the planet. Inside the



The tower for the cancelled Icicap test still stands.

boxes, more than 12,000 high-speed processors crank through incredibly detailed simulations of nuclear-weapons tests. "This is the first machine", East says, "that really gives us button-to-bang capability."

Harnessing the power of supercomputers such as Purple and data from past tests, US weaponeers are working feverishly on an ambitious programme to design a new nuclear warhead that they can certify will work — even without a test explosion. They claim that the new weapon will replace the ageing warheads in the US nuclear stockpile; that it will be safer and more reliable than existing designs; and that it will be easier to build and cheaper to maintain. Some designers informally call it the 'wooden bomb', because theoretically it will be able to sit on the shelf for years with little maintenance. Formally, the new weapon is known as the Reliable Replacement Warhead, or RRW.

## Virtual explosion

For Livermore and its sister facility, Los Alamos National Laboratory in New Mexico, the RRW is the future. It will provide a new generation of weapons designers the chance to work on a nuclear warhead, and give the weapons labs a well-defined project in the post-cold-war era.

NNSA/NEVADA SITE OFFICE

## CONVINCING THE GENERALS

Nuclear weapons scientists may believe that a new warhead, the Reliable Replacement Warhead (RRW) can be built without a test, but they must persuade others in the government if testing is truly to be avoided.

At least some Pentagon officials seem comfortable with the idea of using the RRW without a test. General James Cartwright, head of the US Strategic Command — the umbrella group that oversees the military's nuclear activities — recently expressed support for the programme. "The work that we are doing with the laboratories today would give us reasonable confidence," Cartwright said in an interview published in

June in *Arms Control Today*<sup>5</sup>.

But it remains to be seen whether scientists can convince generals in the Air Force and Navy. In the past, these officers have been enormously resistant to changes to existing weapons, notes Paul Robinson, former director of Sandia National Laboratories in Albuquerque, New Mexico, which maintains non-nuclear components for the warheads.

In the end, opposition often comes down to cash instead of confidence. When changes are made to warheads, the Navy and Air Force must also change their own software and hardware systems for delivering the bomb — all of which costs money.

The National Nuclear Security Administration, which oversees the RRW design competition, says it is requiring the new warhead to be compatible with existing missiles.

For now, the military has not had to pay for the RRW, and the higher-ups at the armed forces aren't paying too much attention to the programme, says one congressional staffer who deals with budgetary issues. But he thinks that as the new designs move forward, it will be increasingly likely that the Pentagon will have to supply funding: "At some point in time, the Department of Defense is going to have to step up to the plate." **G.B.**



A new type of warhead would raise costs for the Navy.

But critics of the programme — including some who designed the current generation of US nuclear warheads — doubt that the RRW can be guaranteed to work without a test. "I just can't believe anyone would prefer a new warhead that's designed by people who have never designed anything before and then

made by people who have never built anything before," says Harold Agnew, former director of Los Alamos. "To me, that's ludicrous."

Despite the criticism, the programme is quietly gaining political momentum. Congressional appropriators, who killed earlier design programmes, gave the RRW project a respectable \$25 million last year. If the programme continues on target, the warhead could enter military service in the next decade.

The debate over the RRW has its roots in the 1992 testing ban, instituted by the former President George Bush as the first step towards signing the Comprehensive Nuclear Test-Ban Treaty. The United States never ratified the treaty, but the government has maintained its voluntary moratorium on testing.

### Historic stockpile

"The test ban symbolizes that the nuclear arms race is over," says Robert Nelson, a physicist and arms-control expert at Princeton University in New Jersey. As long as the United States doesn't test, he says, other nations — including nuclear upstarts such as India and Pakistan — feel enormous pressure to follow suit. And the ban gives the United States a huge advantage over other established nuclear nations, because it already has data from 1,054 nuclear tests. China, by comparison, has conducted only 45.

The end of testing has left the United States with an ageing stockpile of nearly 10,000 nuclear warheads. Most are between 17 and 30 years old, says Nelson, and are of roughly a dozen different designs<sup>1</sup>. All use ordinary explosives to compress nuclear material, often

plutonium-239, which then triggers a series of runaway fission and fusion reactions (see graphic). The warheads aren't easy to maintain because, in addition to normal ageing processes such as rusting, the plutonium in a weapon's trigger, or 'pit' — the component needed to initiate the chain reaction — emits a small but steady stream of radiation. That radiation changes the properties of the plutonium alloy by altering its crystalline structure<sup>2</sup>, which in turn can cause the weapon to fail.

Over the past 14 years, researchers have studied these warheads with a battery of computer simulations and experiments. In recent years, those studies have revealed some significant new details about the old weapons, says

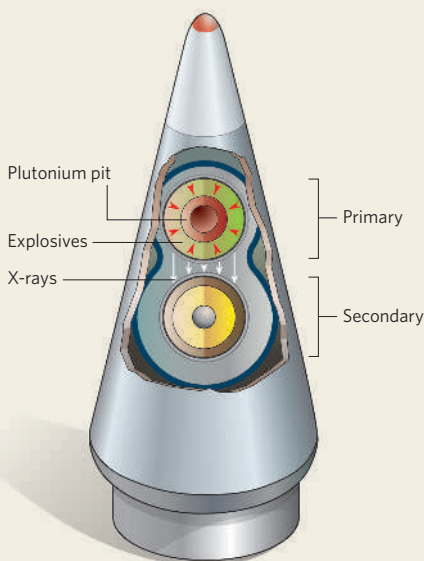
James Peery, who directs the explosive-testing programme at Los Alamos. "They're discovering things and seeing things that they did not expect," says Peery. Details, of course, are entirely classified.

Nobody believes that there are serious problems with the exist-

ing warheads, but the unforeseen discoveries are contributing to anxiety about how long they can be maintained. "I have great reservations that one could use pits that have aged for more than 50 years," says metallurgist Siegfried Hecker, a former director of Los Alamos who is now at Stanford University in California. At some point, Hecker says, the plutonium in the current pits will have to be melted down and remade into new ones.

And that is where the RRW comes in. During the cold war, physicists focused on making the lightest, most explosive weapons possible,

### HOW TO BUILD A NUKE



Modern nuclear warheads consist of two stages: the 'primary' and 'secondary'. For the bomb to work, explosives in the outer shell of the primary must detonate, squeezing a hollow sphere of nuclear material, usually plutonium-239, and triggering a runaway fission reaction. X-rays from the primary then cause atoms in the secondary's fuel to fuse and release still more energy.

**"This is the first machine that gives us button-to-bang capability." — Doug East**



in order to fit many warheads on to a single missile, says Paul Robinson, a retired physicist who spent seven years directing the Los Alamos weapons programme. But in the post-cold-war era, where nuclear showdowns between superpowers seem less likely, missiles often carry fewer warheads, freeing up both space and weight for designers. "Today, you could do a lot to make the things more robust," Robinson says.

### Vision of the future

Weapons designers at Livermore and Los Alamos are now working on RRW designs as replacements for the United States' most abundant nuclear warhead, the W76, which is deployed on submarine-launched missiles. Eight W76s, each destined for a different target, can sit atop a single missile. But today, most missiles routinely carry four.

At the centre of the Los Alamos campus, a gleaming glass-and-steel building houses the lab's alternative to the ageing W76. Past security turnstiles and fingerprint scanners, designers are using Los Alamos's supercomputers — only marginally less powerful than Livermore's Purple — to virtually test their RRW designs. Inside a secured room called 'the cave', 33 projectors display three-dimensional, stereoscopic simulations on the walls and ceiling. Using a joystick, designers can rotate, spin and zoom through each warhead design as it detonates, watching every stage of the explosion.

The simulations are not pure abstractions; they are heavily based on years of experimental data, including those from non-nuclear explosive testing. Both Los Alamos's and Livermore's RRW prototype designs are based on earlier warhead designs that were tested underground, according to one weapons laboratory scientist who requested anonymity. "The designs will be so close that even sceptics



LOCKHEED MARTIN

**Lighter load:** since the end of the cold war, submarine missiles often carry half the warheads they once did.

will accept the simulations," he says.

Few outside the weapons labs know what these simulated alternatives to the W76 look like. But congressional testimony, unclassified laboratory studies, and media reports all point to a number of likely changes.

The most obvious alterations could be made to the weapon's plutonium pit. Adding more plutonium may ensure that the device deto-

nates properly, even after years of sitting on a shelf. Pits could also be redesigned for ease of manufacture, says Hecker. During the cold war, pits were fashioned by shaping sheets of heated plutonium metal — a fast but imprecise technique. "We had very rough specs and then we went and conducted a nuclear test," Hecker says. "As we look to the future, I would definitely vote against doing it that way." He says that the pit for an RRW could instead be cast in a mould.

### A difference of design

Toxic materials in the warhead may also be replaced with more benign substances. Currently, plutonium in the W76 warhead is surrounded by a shell of beryllium, which helps to amplify the initial nuclear explosion. But beryllium is also toxic and carcinogenic, so replacing it with heavier material such as stainless steel would reduce the environmental hazards associated with manufacturing the warhead.

Finally, a report from the weapons labs indicates that designers are considering replacing the lightweight but volatile explosives on the W76's outer shell with a less powerful and more stable explosive called an 'insensitive high

## BRITISH SECRET FORCES?

Is Britain taking part in the Reliable Replacement Warhead (RRW) programme? A recent investigation by *The Sunday Times* claimed that British nuclear scientists "have been secretly working with Americans on a replacement".

Ministry of Defence officials deny any British involvement in the project. "The article contains a number of inaccuracies and misleading statements,"

notes Matthew Willey, a spokesman for the ministry. In particular, a joint US-UK test cited as being part of the new RRW programme was, according to Willey and US officials, just a routine test of an existing, ageing warhead.

That isn't to say that UK weapons scientists aren't following what their US counterparts are doing. Britain's nuclear deterrent consists entirely of submarine-launched, W76-style warheads — the kind

now being considered for replacement by the RRW programme.

British weapons scientists are trying to determine what to do about these weapons, says David Overskei, a California-based weapons consultant who chaired a US panel last year that examined the RRW and the US nuclear complex. "As far as I know, they are not involved with the RRW," he says. "But they are keenly, keenly interested." **G.B.**



explosive'. This would increase the weapon's size and weight but decrease the likelihood of an accidental detonation during storage.

But will these changes really lead to cheaper warheads without the need for testing? A dozen current and former designers unanimously agree that changes might simplify the process of maintaining warheads. But there is far less accord on whether the new warheads would require testing, or whether they would be affordable when compared with remanufacturing the existing stockpile.

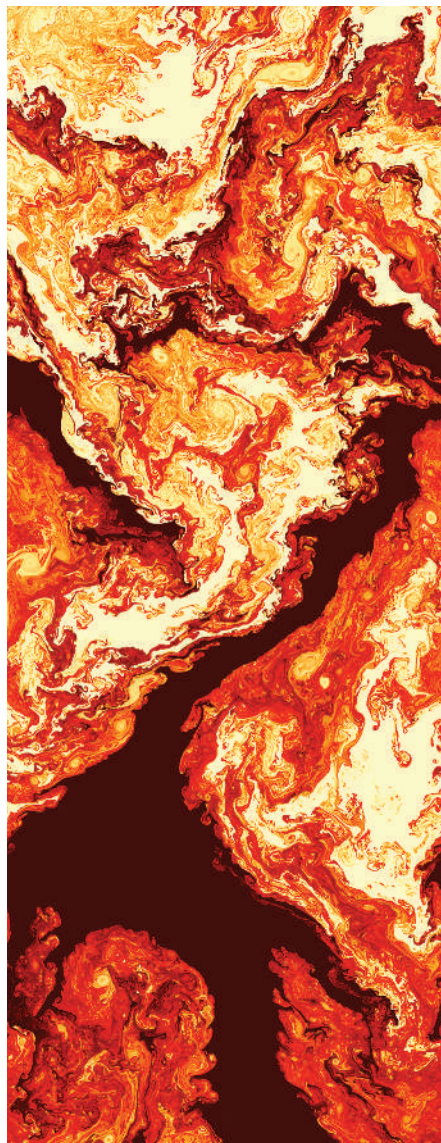
Agnew, who oversaw Los Alamos during the design of the W76, is among the most vociferous critics of the RRW programme. Very little, he argues, is known about how alterations affect performance. "These nouveau designers don't know what the margin is. In fact, no one knows what the margin is," he says. For example, adding more plutonium to a warhead's pit doesn't necessarily make it more reliable, he argues. It could instead make the warhead more likely to accidentally explode, or it could overheat the 'secondary' fuel which produces most of the weapon's power. There's simply no way to tell without a test, he asserts. "If you really believe that the nuclear deterrent is important," he says, "you shouldn't put things in the stockpile that aren't tested."

Sidney Drell, a Stanford University physicist, and others also question whether the changes are much of an improvement. A 1994 study by Drell and Robert Peurifoy, formerly at Sandia National Laboratories in Albuquerque, New Mexico, showed that rocket fuel, not high explosives, would be the most likely cause of an accidental explosion<sup>3</sup>. As a result, the Navy decided not to use insensitive high explosives on the W76.

### Where doves meet hawks

And then there is the question of cost. The programme to look after the existing US nuclear stockpile is expensive — more than \$1.4 billion a year. Members of Congress are interested in the RRW because of claims that it could save money, even though the labs have not released a total cost estimate. But Richard Garwin, a former bomb designer, argues that the reason the stockpile costs so much has more to do with its size than with its age. More money could be saved by cutting the number of existing weapons from 10,000 to 2,000, he says. With a diminished arsenal, says Garwin, "we will be able to maintain current weapons indefinitely."

Not all former weapons designers are so critical of the RRW programme. Herbert York,



Computer simulations of turbulent mixing are relevant to modelling how warheads might perform.

Livermore's first director, believes that early warhead designs, which have already been tested, could provide a reliable basis for designing a much simpler but much heavier warhead. "I'm not sure it would be practical, but I believe they could be designed to be stockpiled without testing," he says.

Despite the criticism, Republicans and Democrats are looking favourably on the programme. Hawks like the programme because it will allow the United States to train a fresh generation of weapons designers. And doves, who have torpedoed earlier weapons programmes<sup>4</sup>, are wooed by the claim that the RRW will not need to be tested. Congress is

likely to more than double the programme's budget for next year. And the military has lent its tentative support to the project (see 'Convincing the generals').

All this is good news for the country's ageing nuclear weapons complex, according to long-time observers. "The US nuclear programme has been in a cul-de-sac since the end of the cold war," says John Foster, a former director of Livermore who is chairing a panel on the complex for the Pentagon's Defense Science Board. Since the collapse of the Soviet Union, lab morale has sagged and efforts to refurbish warheads have fallen behind schedule, Foster says. The RRW programme provides the labs with a fresh challenge and clear vision. "The RRW," he says, "would catalyse the enterprise from design through production."

And, indeed, the programme does seem to have a reinvigorating effect (see 'British secret forces?'). Weapons designers are thrilled to be working on their first new warhead in two decades. Earlier this spring, the Livermore team ran a huge simulation of its RRW design. At Los Alamos, scientists are about to conduct a non-nuclear explosive test to check some of their calculations.

Both teams have submitted their designs to a review committee, where they are being peer-reviewed. One of the two designs will be selected as the basis for a development programme later this year. A second competition may even be held in 2007, for designs for another RRW. If all goes well, pits for the first warhead could be manufactured as early as 2012.

For now, the Reliable Replacement Warhead remains a series of zeros and ones, in the huge supercomputers at Los Alamos and Livermore. But back in the Nevada desert, the structure that once housed Icecap still looms above the Joshua trees. The rigging to hold a bomb — all 225,000 kilograms of it — remains in place, along with hundreds of metres of copper cable designed to carry data signals a few nanoseconds ahead of the blast.

Icecap, in short, is ready for the next US underground test. "Should we come back to nuclear testing?" Ernie Williams cheerfully tells his visitors, "it seems reasonable we'd start with this one."

**Geoff Brumfiel is Nature's physical sciences correspondent in Washington DC. See Leader on page 2.**

1. Norris, R. S. & Kristensen, H. M. *Bull. Atom. Sci.* **62**, 68–71 (2006).
2. Hecker, S. S. *MRS Bulletin* **26**, 672 (2001).
3. Drell, S. & Peurifoy, B. *Annu. Rev. Nucl. Part. Sci.* **44**, 285–327 (1994).
4. *Nature* **438**, 139 (2005).
5. Boese, W. & Pomper, M. A. *Arms Control Today* [www.armscontrol.org/act/2006\\_06/CartwrightInterview.asp](http://www.armscontrol.org/act/2006_06/CartwrightInterview.asp) (2006).

**"These nouveau designers don't know what the margin is." — Harold Agnew**



# HARVEST OF HOPE

An economist believes that a five-year aid effort in a dozen villages across Africa can teach the world how to defeat poverty. **Sarah Tomlin** reports on the project's progress in Rwanda.

Hyacinthe Mukaritaganda of Kagenge village helped build a communal water tank, as part of the Millennium Villages project.

Celestin Ndahayo smiles broadly at me from below the corn (maize) that towers a metre or more above him, his daughter Annalita clutching his hand. This is the first corn harvest he has seen here in almost ten years. There was a smaller harvest of beans and sorghum in 2001; last year there was nothing. In the years without good rains, the people of Kagenge (sometimes called Mayange) in Rwanda survive the best they can. Some walk four nights and three days to reach a more productive region. Ndahayo sometimes takes construction jobs to support his wife and four children.

Hyacinthe Mukaritaganda's husband is one of those currently working elsewhere, leaving her to manage their land and look after three children on her own. This season she planted corn on one-fifth of their land, a short walk from Ndahayo's homestead. Thanks to the rains, she is expecting a good harvest, which should provide enough seeds to plant all 2.5 hectares next year. Then, she hopes, her husband will stay at home to help.

The rains, though, are not the only things bringing hope to Kagenge. In 2005, the village was chosen to take part in the Millennium Villages project. Led by the Earth Institute at Columbia University in New York, the project is applying a range of poverty-slashing interventions to 12 sites across Africa (see map). The idea is not just to show that interventions in a number of different areas, properly coordinated and financed, can make a sustainable change to the lives of the world's poorest communities. It is to show how that can be done quickly in a way that can be replicated easily.

Donald Ndahiro, an agronomist trained in Uganda, is the project's agriculture coordinator for Rwanda. He says that when he arrived in Kagenge late last year conditions were desperate. "The villagers were emaciated." They wanted food aid more than they wanted the agricultural advice, drought-resistant seeds, fertilizer and new techniques that the project was offering. "They thought we were making fun of them," Ndahiro says. "We were telling them how to plant, how to harvest, but they were saying they were never getting any good rains. We told them to get organized."

Five months later and the villagers are getting organized. Ndahayo is a member of the agriculture committee that will decide what to do with the surplus from this year's corn harvest. Mukaritaganda is helping to clear land for a tree nursery (villagers sometimes walk

S. TOMLIN







Fertilizer, seeds and advice are being given to 12 villages in Africa, to demonstrate how properly coordinated interventions can make a sustainable difference to people's lives.

ten kilometres to gather firewood) and was part of the team that just built a communal tank to collect rainwater. She invites me with pride to a ceremony in which certificates are awarded to her and the 25 other villagers who worked on the tank.

### Leading the way

The Millennium Villages project aims to provide improved resources and techniques not only in agriculture, but also in health, education, transport, energy and water provision, and financial management. The plan is

to achieve the United Nations' Millennium Development Goals (see above, right) for the 5,000 or so people in Kagenge, and for the tens of thousands of people in the 11 villages elsewhere within 5 years — 5 years ahead of the UN target date.

The eight goals, committed to by 189 heads of state in 2000, include halving the number of people living on less than US\$1 a day and controlling malaria by 2015. Progress so far has been limited, especially in Africa — far too slow for the impatient economist Jeffrey Sachs, head of the UN Millennium Project and the Earth Institute. Sachs wants the 'research villages' and the data that they provide to offer ways of picking up the pace: "The idea is to demonstrate a practical path and to mobilize governments."

The man in charge of making such a demonstration is Josh Ruxin, a Columbia University public-health expert and the project's director in Rwanda. Ruxin, imbued with an impressive energy and passion, was initially sceptical of the village-by-village approach: he wanted to target a millennium country not an isolated village. But Ruxin is encouraged by the Rwandan government's own ambitious poverty-reduction strategy, known as Vision 2020.

Ben Karenzi, the Rwandan health ministry's secretary general

### MILLENNIUM DEVELOPMENT GOALS

- Eradicate extreme poverty and hunger
- Achieve universal primary education
- Promote gender equality and empower women
- Reduce child mortality
- Improve maternal health
- Combat HIV and AIDS, malaria and other diseases
- Ensure environmental sustainability
- Develop a global partnership for development

says "We believe it's possible, especially with the focused leadership we have and the commitment of our people, to make Rwanda a mid-level income country by 2020." In the context of that commitment, Ruxin is confident that with the help of the Millennium Villages project, Rwandans can succeed in not just turning round one village, but in transforming life for poor farmers across the country.

### Money cares

Part of Ruxin's confidence comes from an assessment of the government. In the aftermath of the genocide of 1994 and the resettlement of some two million returnees from neighbouring countries, 64% of Rwanda's population was living in poverty (on less than US\$1 a day) in 2000. But despite its internationally criticized role in the Congo war, the government of Paul Kagame is widely seen as committed to poverty reduction, and as embodying principles of good governance from the top down (for example, all ministers are required to declare their annual income).

Ruxin believes that good governance will be an important factor in the long-term success of the millennium villages. Those running the project have deliberately avoided what they see as the worst African regimes. But they say that even in corruption-prone nations, such as Ethiopia and Kenya, the research villages so far remain free of corruption. Sachs points out that if you focus on supplying commodities, such as seeds, fertilizer and nets for protection against malaria, "there's very little money that changes hands". That said, Sachs is less worried than many about corruption; he knows people criticize this lack of concern, but doesn't care. "Corruption is way down the list of practical issues," he argues, Africa's miserable roads, poor soil and endemic disease burden are at the top.

Indeed, the road from Kigali, Rwanda's capital, to Kagenge in the Nyamata district, throws up choking red dust in the dry season and can be impassable in the rainy season. Although the land looks green from the air, the rains can be infrequent and Ndahiro confirms that the soils are poor. Some 70% of the patients at the village's clinic have malaria and the district has one of the country's highest levels of HIV, at about 13%.

As well as suffering from Sachs's top three





S. TOMLIN

problems, Kagenge has its own particular sadnesses. The local mayor, Gaspard Musonera, lost three-quarters of his family in the 1994 genocide. "Nyamata district lost more than half of its population," he says. "The implications and consequences of that you can imagine for yourself."

Kagenge itself is a community created since the genocide. Half the households live in settlement housing — or *umudugudu* — built by the government

for survivors and returnees. Ndahiro is himself a returnee, living in Nyamata near the church where 10,000 people were murdered in 1994 — the blood stains on the walls and altar cloth remain as a memorial. He recalls how lifeless the town was when he arrived in 1997. People were bitter, he says; some didn't want to continue living.

### Grand plan

Musonera sees the Millennium Villages project as a sign of hope for the most vulnerable people in his district, and a big test for poverty-reduction measures. "If it can be done here, it means it can be done elsewhere," he says — and that indeed is the point. The project is not just about breaking the cycle of poverty in 12 villages, but about learning how to do it in 1,200 or 12,000. Sachs's plan is to show that with a five-year investment of about US\$550 per person — \$50 a year from the project, \$30 from government, \$20 from other donors and \$10 from the villagers — an integrated package of low-cost interventions can produce long-term financial sustainability in a way that not only can be repeated but can also be scaled up. The project plans to grow to 78 villages this year by creating clusters around the 12 original research villages. The expansion is being funded by the US Millennium Promise charity, which has so far raised \$100 million to support Sachs's vision.

Not everyone is convinced that the Millennium Villages project will succeed. Ecologist Ian Scoones at the Institute of Development Studies at the University of Sussex in Brighton, UK, is a member of the Future Agricultures Consortium, which was put together by the UK Department for International Development to focus on African agriculture and development. Scoones points to the Integrated Rural Development and 'villagization' schemes that tried to boost African agriculture in the



Angelique Kanyange is one of a handful of doctors working in rural Rwanda.

1970s and 1980s. "They created little islands of success but when donors pulled the plug they all collapsed." Scoones says he is very pleased that the millennium villages are putting African agriculture back on the map, but he is afraid of old mistakes being repeated, and worried about things moving too quickly. "India launched its green revolution in the mid-1960s on the back of decades of solid investment and research," he points out. "It didn't

happen overnight."

For his part, Sachs sees patience, like a well-developed sensitivity to the issue of corruption, as an overrated virtue. He has no worries about moving too fast. "It happens to be an emergency," he says. And he has no illusions about the projects working as examples simply by word of mouth. "This is not viral. You can't do it without resources," Sachs notes — as ambitions grow, so must spending. The biggest risk, he says, is for official donors to sit on their hands.

The goal is not to do without large transfers of money to Africa, it is to work out how to make those transfers more effective. After all, the individual interventions used in the Millennium Villages project are tried and tested methods, even if they haven't been applied all together in one location before. Asked about the target of reaching the millennium goals in five years, Celina Schocken, an international-affairs fellow at the US Council on Foreign Relations, says "I absolutely believe they will succeed. I don't see how they can't."

But she's less convinced about how scale up will be achieved. "What good is an island of prosperity anyway?" she asks. Scoones agrees that the big question is: "How, without that external support, do you replicate?"

That is the question Sachs, Ruxin and their colleagues are trying to answer. By documenting all the inputs and outputs for each research village they hope to tease out the synergies between overlapping interventions. Measuring 35 indicators for the 8 goals across several hundred households in 12 villages is time consuming and costly, but it is necessary to show not just that the investments work, but also how they work, and how they can work better. Only then can they be scaled up to the truly monumental level envisioned by Sachs, who wants to see development aid change the course of history. "I think the

**"The idea is to demonstrate a practical path and to mobilize governments." — Jeffrey Sachs**



Donald Ndahiro, an agronomist working for the Millennium Villages project, has helped to transform conditions in Kagenge in just five months.

biggest challenge is the defeatist attitude of the official donor community," says Sachs. Such rhetoric reinforces the suspicion that Sachs is unwilling to learn from lessons of the past. "People in the development community see some benefit in the publicity Jeff Sachs gets," says Schocken, who used to work with Ruxin in Rwanda, "but they've seen these ideas before."

### Skill shortage

In Kagenge, the villagers assembled for the water-tank certificate ceremony are briefly reminded of the international debate over their future. "This is an important day for the project," Ruxin tells them during a short address, "You are now the teachers for us and for the world." Some of the farmers I met in the fields yesterday have donned suits and ties for the occasion. Each villager who received training from visiting Kenyan water specialists receives a signed certificate — the expectation is that they will take the skills they have learned and pass them on to others. After many more speeches by village leaders, the villagers distribute soft drinks and, for those who can stomach it, fermented sorghum, the local brew.

In the weeks before the corn is harvested,



the contrast between Kagenge and the surrounding area is already striking. An emergency feeding centre supported by the UN Children's Fund UNICEF and the World Food Programme was set up in Kagenge in early March, in response to reports of serious malnutrition following last year's drought. Four hundred people from the wider local population are still receiving weekly rations, but not those of Kagenge. The bean crop and corn picked straight from the fields before the harvest mean that they have enough to eat.

They also have a functioning health centre, which serves Kagenge and four neighbouring communities of similar size. The centre now has its own doctor, Angelique Kanyange, known to everyone as Dr Angelique, and its nursing staff has doubled in number. Dr Angelique is zealously improving the nurses' cleaning procedures with demonstrations of the use of a broom and disinfectant. Today, the clinic is seeing more than 35 patients a day, as well as some 50 mothers bringing children for immunization. One of the new patients is Musabyimana, an 8-year-old boy who is blind and in pain because of severe cataracts. His mother noticed his poor sight when he was three months old, but this is his first visit to the clinic. Dr Angelique is not sure what caused the cataracts, but there is hope, she says, because Musabyimana seems to be able to

detect some light and colour. She will refer him to a specialist for treatment. For now, the project will fund it; the mother, a widow, could never afford it.

Rural parts of Rwanda have community medical insurance schemes, but only 12% of families in Kagenge have cover. The goal of the Millennium Villages project is to get 100% coverage, with the hope that as the clinic becomes more useful to patients, more will join the scheme.

### From village to province

It is here, though, that scaling up looks harder than it does in agriculture. Buying more fertilizers is easier than making more doctors. "Angeliques are hard to come by," admits Ruxin. Indeed, Dr Angelique is the first government-appointed doctor in a rural health centre, demonstrating the government's commitment to the Millennium Project but also, perhaps, the project's weakness. "The president of Rwanda says 'I want this village to work', so they are going to get the best," says Schocken. There are currently about 200 doctors in the country. The medical schools may be able to produce 60 or 80 a year, but the country has a long way to go to reach the World Health Organization's minimum recommended level of one doctor per 5,000 people.

Sachs claims recruitment problems can be overcome with decent salaries. Although the project is mostly about spending money on physical resources, he is in favour of top-up payments for doctors. But even with targeted salary increases, a country such as Rwanda suffers skill shortages in every sector.

Kagenge currently has a team of ten dedicated people who work long hours to motivate the villagers and document their progress. Detailed accounts were not made available to *Nature*, but in its first year the Kagenge project will spend as



**"We want a millennium province, not just a millennium village."**  
— Theoneste Mutsindashyaka

much on personnel as on materials. The budget for the cluster villages being set up in addition to the original 12 is smaller, and they will have much fewer support staff. "Research on top is extremely expensive," notes Ruxin, explaining that in future, and in villages that aren't research focused, costs should be much lower. But Sachs's claim that "the science behind this is broadly transferable without needing large teams" has yet to be put to hard tests.

The budgets matter to Theoneste Mutsindashyaka, a former mayor of Kigali and the governor of Rwanda's eastern province, which includes Nyamata and covers a quarter of Rwanda's population. He is a great fan of the Kagenge project, in part because it fits so well with the government's Vision 2020. He wants the figures so that he can roll out projects informed by the experience more widely. "The documentation is very important to me because I have to negotiate with partners," he says. He is impatient to get moving on the next stage of the project: "We want a millennium province, not just a millennium village," he says.

Within the next year, Mutsindashyaka wants to set up a Kagenge-like village in each of his provinces' seven districts. "We are going to move village to sector, sector to district, but you have to have money," he says. And he is certain he can sell the idea to his friends all over the world, from Quincy Jones to Donald Kaberuka, the president of the African Development Bank. And although scaling up to 3,600 villages is daunting, the governor says he only needs the numbers from Kagenge to get started: "I am total 100% confident that the project will succeed."

From Sachs to the president to the governor to the mayor, the ambitions for transforming the country are vast. But in Kagenge, despite the good rains, the villagers themselves remain wary. They are not as confident that they will achieve rapid progress as the project leaders. Anxiety about what to do with the harvest surplus is high. Celestin Ndahayo and other farmers worry about whether they can really afford both to sell corn and store enough for food security; they are not sure they believe Ndahiro's forecasts for the yields of their smallholdings. And what if the rains don't come next year? In his experience, says one *umudugudu* farmer, when a project is here, then the rains come. Back in 2001, an organization helped them to plant cassava and sweet potato and the rains came. But when they left the rains stopped. So as long as the Millennium Villages project is here he believes it will rain again. He doesn't believe, yet, that his village can learn to flourish in the project's absence. ■

**Sarah Tomlin is a senior News Feature and Commentary editor.**

D. NDAHIRO



## Reviewers' reports should in turn be peer reviewed

SIR — As a young investigator with limited experience in the peer-review process, I would like to add a perhaps naive suggestion to the comments made by Rory Wilson (*Nature* **441**, 812; 2006) and others in Correspondence and at [www.nature.com/nature/peerreview/debate/index.html](http://www.nature.com/nature/peerreview/debate/index.html).

As a result of my conversations with peers and mentors, I suggest that, when assessments are complete, journals ask reviewers to review the other reviewers' comments before the editor makes a decision about publication. Although this assessment is traditionally reserved for the editor, by holding reviewers accountable, thereby encouraging fair and reasonable reports, editors will be better able to assess the suitability of a recommended rejection, revision or acceptance. I believe my suggestion would be particularly useful in journals for which the editor is a practising scientist. In these cases the editor is not anonymous, so may not assess reports as freely as reviewers who have this protection.

Although the time taken for the initial review process may be increased by my suggestion, I believe it would result in a fairer process, as the editor would benefit from the feedback in the decision-making stage and the reviewers would be given an incentive to provide their services fairly.

**Alexandra List**

Department of Psychology, University of California, Berkeley, California 94720-1650, USA

See *Nature's* web debate on peer review at [www.nature.com/nature/peerreview/debate/index.html](http://www.nature.com/nature/peerreview/debate/index.html)

## Judge a paper on its own merits, not its journal's

SIR — Your News story "Cash for papers: putting a premium on publication" (*Nature* **441**, 792; 2006) reports that South Korea is joining China and Pakistan in rewarding researchers with cash for publications in elite journals. Presumably the higher the journal's impact factor, the more valuable the reward.

The impact factor may be a good indicator of the quality of a journal, but it is misleading to judge an individual publication by the name of the journal in which it appeared. A journal's impact factor, as defined by the ISI, is the average number of citations per article the journal received during a given period. Isn't the number of citations a paper receives a better, more direct measure than the average of papers in the same journal?

A journal is good because it contains many high-impact papers, but not all papers in that journal necessarily have high impact. Papers that are published in low-profile journals but receive an equally high number of citations

should not only be rewarded, but be rewarded more. In judging an individual publication, what counts should be its real merits, not its batch label.

**Shu-Dong Zhang**

MRC Toxicology Unit, Hodgkin Building, Lancaster Road, Leicester LE1 9HN, UK

## Unpredictable Sun leaves researchers in the dark

SIR — We agree with the view expressed in your News Feature "The dark side of the Sun" (*Nature* **441**, 402–404; 2006) that solar magnetic activity is important. It leads to such dangerous events as solar flares and coronal mass ejections, as well as being associated with fascinating phenomena such as sunspots and the solar cycle.

The Sun's magnetic field is now believed to be generated by a hydromagnetic dynamo acting deep within its interior, where streams of highly ionized plasma generate electric currents and, in turn, magnetic fields. Progress in dynamo theory is extremely difficult, as it can be made only by understanding the interaction of turbulent plasma motions with magnetic fields. Indeed, the extreme conditions within the solar interior make this a formidable task; understanding even the much less turbulent environment of the Earth's atmosphere stretches current theories to their limits, and weather prediction is notoriously precarious.

It is in this context that current attempts to predict magnetic cycles on the Sun must be viewed. The model proposed by Mausumi Dikpati and her team, highlighted in your News Feature, relies on parametrization of many poorly understood effects. Although such parametrized models have been widely (and legitimately) used to explore specific features of dynamo processes, they have no detailed predictive power. Indeed, there is vociferous debate in the field, not just about the size of many of the effects included in Dikpati's (and many other people's) models but even their signs. Moreover, the dynamo equations are extremely nonlinear; the solar dynamo is believed to exist in a state of deterministic chaos, making prediction intrinsically yet more difficult. Any predictions made with such models should be treated with extreme caution (or perhaps disregarded), as they lack solid physical underpinnings.

Of course it is interesting to speculate on what direction solar magnetic activity might take in the future. Recent sunspot cycles have been exceptionally vigorous, as noted by S. K. Solanki and colleagues (*Nature* **431**, 1084–1087; 2004). It is well known that, in the past, such episodes of high activity have tended to be followed by a dramatic crash

into periods of severely reduced magnetic activity, termed Grand Minima. Although we would not presume to predict that this will happen soon, it would certainly be interesting to witness such a collapse.

**Steven Tobias\*, David Hughes\*, Nigel Weiss†**

\*Department of Applied Mathematics, University of Leeds, Leeds LS2 9JT, UK

†Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 0WA, UK

## Second thoughts on who goes where in author lists

SIR — I thought I understood the guidelines for determining scientific authorship: the individual making the greatest intellectual contribution is the lead author, followed sequentially by those making progressively lesser contributions. In addition, the final-author slot is sometimes reserved for a lab head or project initiator, who may have made little direct contribution to the paper but deserves some vague honour nonetheless.

But now I am confused. A collaborator of mine at the University of Cambridge asked to be moved from second to last position on a four-authored paper. When I asked why, he said the British Research Assessment Exercise (RAE), which determines departmental rankings and government funding, gives greater credit to the final than even the second author on a multi-author paper. My confusion deepened when two other colleagues — both Americans — had a lively disagreement about who would be last author on a paper with seven authors.

The 'communicating' (or 'corresponding') author is often, but not always, the lead author. If he or she is not the lead, is some special significance attached to this? Does it count for something on the RAE? Some disciplines have evolved their own idiosyncratic rules. I have also noted that another common convention is to list authors alphabetically, but does the RAE know if Williams made a lesser or greater contribution than Anderson?

Is there a set of coherent authorship rules written down somewhere that, in my 20-year research career, I have managed to miss? If not, then perhaps there should be.

Please note that I am the first, last and communicating author on this Correspondence.

**William F. Laurance**

Smithsonian Tropical Research Institute, Apartado 2072, Balboa, Panama

*Nature's* most recent Editorial stating its policy on this perennial topic, "Games people play with authors' names" (*Nature* **387**, 831; 1997), can be seen at [www.nature.com/nature/journal/v387/n6636/full/387831a0\\_fs.html](http://www.nature.com/nature/journal/v387/n6636/full/387831a0_fs.html) — Editor, *Nature*.

## BOOKS &amp; ARTS

# Changing our minds

Our environment can affect the way our minds develop, but the relationship is complex.

## Brain and Culture: Neurobiology, Ideology, and Social Change

by Bruce E. Wexler

Bradford Books: 2006. 320 pp. \$34, £21.95

### Paul Bloom

Perhaps the mind begins as a blank slate and we start off, to use Rousseau's phrase, as "perfect idiots". At the other extreme, it could be like a Swiss-army knife, a collection of innately structured neural modules. Perhaps the mind starts off being modular and becomes flexible through development — or perhaps it starts off undifferentiated and becomes modular. Maybe what gives us our uniquely human mental powers is the capacity for complex language. Or cultural learning. Or meta-representation. There is no shortage of one-line theories of human nature.

In his engaging new book, *Brain and Culture*, Bruce Wexler argues that what is interesting about the mind is neither our inborn nature nor the structure of our environment — it is how the two interact. There is, he explains, a principle of internal–external consonance: humans are driven to match their internal neurological structures to the external environment.

In the first half of the book, Wexler discusses the developmental implications of his theory, arguing for a process in which children's neural structures are moulded and transformed by the external environment — it is our nature to be nurtured. Although he concedes that some aspects of human nature might be inborn, his sympathies lie with Rousseau. There is a nicely provocative passage in which he compares the brain and the stomach, suggesting that the stomach is, in some sense, smarter. The stomach can work independently of the environment, whereas the brain cannot: "The brain recreates in itself a representation of environment input which, especially in the formative years, conforms highly to the complexities of that input."

The second half of the book explores the flip side of the consonance principle: once neuroplasticity is reduced in late adolescence, we stop changing our minds to fit the world, and instead try to change the world to fit our minds. We prefer familiar things and people; we reject new ideas; and we become miserable and ill when faced with changing circumstances, as when a loved one dies or when migrating to a new



J. GREENBERG/ALAMY

Children adapt: a boy learns to communicate with his deaf friend by sign language.

country. There is little neuroscience here; Wexler instead adroitly makes his case by drawing on evidence from fields such as social psychology, history and political science.

*Brain and Culture* is a deep, thoughtful and intellectually ambitious book with a high ratio of ideas to pages. It is also gracefully written, very clear and accessible. But the main argument — that there is a consonance principle — is not persuasive.

Consider language. As a staunch believer in the power of the environment, Wexler says that language is a property of cultures, not of brains, and concludes that children who are not exposed to language would never learn to talk. This is a sensible enough prediction, but it seems to be false. Several studies, by Susan Goldin-Meadow, Ann Senghas and others, found that deaf children who are not exposed to a sign language will often create a language themselves. Even in normal language development, children go beyond the input, quickly developing an abstract and generative appreciation of vocabulary and syntax that enables them to produce and understand sentences that they have never before heard.

Elsewhere, Wexler stresses the importance of parents, drawing on the psychoanalytic literature to argue that the social environment of a caregiver plays a powerful role in shaping the child's mind. This is an important point;

some nurturing is plainly essential for the normal development of primates, including humans. On the other hand, one of the striking findings of behavioural genetics is that individual differences in intelligence, personality and temperament have little to do with how children are raised. Instead, roughly half the variation is due to genes, and other half to non-shared environment — that is, environmental factors that are independent of how parents treat their children.

Is the consonance principle true of adults? It is easy to find examples where we seek out the familiar and dislike the new, but it is just as easy to find cases that work the other way round. Studies of happiness find that new experiences delight us, but that these experiences lose their effect as they grow familiar. This is why some people are driven to continually seek out novelty, something that happiness scholars have dubbed the 'hedonic treadmill'.

Wexler describes how the death of a loved one usually causes grief and a period of mourning, and interprets this as illustrating the consonance principle at work, as it "vividly reveals the effects of an abrupt disjunction between internal structure and external stimulation, and the time and effort necessary to recreate a comfortable consonance". But a more plausible explanation is that the



bereavement is due to a specific sort of disjunction — the loss of someone you love. After all, falling in love is also an abrupt disjunction, but it is often a lot of fun.

Given the range of adaptive problems that humans face (from both an evolutionary and development perspective), there is no reason to expect a single principle that governs interactions between the mind and the environment. In some domains, developmental

malleability makes sense; in others, it does not. Sometimes adults should hate the new, when a loved one dies, for example; sometimes we should embrace it, such as when starting a promising relationship. The relationship between the mind and the environment is too complex for a one-line theory. ■

Paul Bloom is in the Department of Psychology, Yale University, New Haven, Connecticut 06520-8205, USA.

experiments and smart approximations such as prized virtues in fluid dynamical research.

Until now, anyone interested in the history of this subject has usually had to turn to books such as *History of Hydraulics* (Institute of Hydraulic Research, 1957) by Rouse Hunter and Simon Ince or to John Anderson's *A History of Aerodynamics* (Cambridge University Press, 1997), both strongly oriented to specific engineering disciplines, and a few other rather specialized works. Oliver Darrigol's *Worlds of Flow* is the first book to see hydrodynamics in the wider context of the history of ideas in science. The subject has finally found the distinguished historian it deserves, and the serious history it demands.

Darrigol approaches the subject through the evolution of the concepts that now describe the motion of fluids — viscosity, vorticity, waves, instability, turbulence. He begins with what he calls the small elite of eighteenth-century Swiss and French 'geometers' (including Daniel and Johann Bernoulli), and progresses to the engineers, mathematicians and physicists of the 19th century. During this time, the subject was divided into the ideal world of the hydrodynamicist, who did beautiful mathematics that often failed reality checks, and the real world of the 'hydraulician' who collected useful formulas disconnected from dynamics. These two 'worlds' of flow evolved separately, generally with scorn for each other.

But around the end of the nineteenth century and into the twentieth, many engineers began to examine the foundations of the subject in their own rather pragmatic ways: Osborne Reynolds's studies of turbulent flow, William Rankine's analysis of shock waves and Ludwig Prandtl's many distinctive theories came to characterize an emerging 'engineering

## A turbulent history

### **Worlds of Flow: A History of Hydrodynamics from the Bernoullis to Prandtl**

by Olivier Darrigol

Oxford University Press: 2005. 376 pp. £35, \$74.50

### **Roddam Narasimha**

The continuing fascination of hydrodynamics — or its modern, more inclusive offspring fluid dynamics — is due to the fact that many phenomena (such as turbulent flows) that we can observe with our unaided senses pose deep scientific problems that have not been solved to this day. Those unaided observations have led artists and scientists to wonder at the beauty, majesty and waywardness of flows over the centuries. Leonardo da Vinci's pictures of vortices, Hokusai's prints of waves, and the unknown Sanskrit poet's celebration of the splendid diversity of flowing water in current, wave, foam and spray — all these are matched by the scientist's struggle to understand flow and the engineer's attempts to manage it.

The governing equations of hydrodynamics were first written down by the French engineer Claude Louis Navier in 1822. Those equations (with which the name of George Stokes is also associated) remain valid, so the fact that turbulence, for example, has resisted a final solution must be attributed to the inadequacy of our mathematics to handle the strong nonlinearity of the equations and the almost universal tendency of flows to crumple into one form of instability or other except under the mildest conditions. John von Neumann saw the problem clearly when he said that "The impact of an adequate theory of turbulence on certain very important parts of pure mathematics may be even greater" (than on fluid dynamics). The basic mathematical nature of the problem is now being more widely recognized: one of the seven million-dollar 'Millennium Prize' problems identified by the Clay Mathematics Institute in Cambridge, Massachusetts, has to do with Navier–Stokes solutions. Understandably, it is this very inadequacy of the mathematics that has made physical insight, clever



**Fluid power:**  
*The Great Wave* by the seventeenth-century Japanese artist Hokusai.

science'. All of these were inspired by the earlier work of physicists and mathematicians, and often appealed to the Navier–Stokes equations. But they also introduced novel approximations and did not hesitate to use purely numerical methods when no analytical solutions were available. Meanwhile, the hydrodynamicists discovered that the real-life problems that the engineers were tackling involved deep questions in physics and mathematics. The study of the instability of fluid motions — the subject of Darrigol's fifth chapter — exemplifies these developments best: the two cultures began to meet, as the formulations of William Orr, Arnold Sommerfeld and Lord Rayleigh led both to the striking work of G. I. Taylor on cylindrical Couette flow and Werner Heisenberg on plane Poiseuille flow,

and to the prescient work on boundary layers by Prandtl and Walter Tollmien.

Darrigol handles these developments with scholarship, insight and charm. One of the fascinating episodes he describes is the vortex theory of matter. For William Thomson (later Lord Kelvin) the idea that matter was made up of vortex rings had an extraordinary appeal: but the theory needed vortex rings to be stable, which unfortunately they were not. It seems to me that it was such failures, amid the striking success of James Maxwell's electromagnetic theory and other exciting developments in relativity and quantum mechanics, that were largely responsible for physicists' fading interest in fluid dynamics. It was thus largely left to the applied mathematicians and the engineers to fashion whatever successes the

twentieth century could claim.

Darrigol takes us through these developments in the still incomplete history of the subject with a rare balance that accurately reflects its rich complexity (although I did miss mention of V. W. Ekman's friction layer in the rotating ocean and L. F. Richardson's heroic failure at numerical weather prediction). This is a book that all practising fluid dynamicists must read: I hope there will be a paperback edition soon, so that the strange history of the subject that Darrigol describes with such insight will become part of the intellectual legacy of interested students in engineering, mathematics and physics. ■

Roddam Narasimha is at the Jawaharlal Nehru Centre for Advanced Scientific Research, Jakkur, Bangalore 560 064, India.

## Form and flow

A sculptural approach to the heart.

**Philip J. Kilner**

During my medical training and early years as a doctor, I was repeatedly drawn to art, even if it seemed to me then to lack relevance to medical work. Art felt liberating and satisfying, but subjective and impractical, whereas the kind of scientific thought that I was coming to adopt seemed convincing and applicable, yet strangely out of touch with human feelings or values. I left medicine for several years and studied sculpture at Emerson College in Sussex, England.

As Martin Kemp has pointed out, science and art are not homogeneous categories (*Nature* **434**, 308–309; 2005). What I found at Emerson College was an observational approach to natural science, and an artistic approach that was nourished by natural forms and processes without necessarily being representational. This was a rewarding mix, fostering discovery, creativity and a sense of wonder. The receptive, phenomenological approach we used owed much to Johann Wolfgang von Goethe (1749–1832), who regarded his scientific studies as more significant than his literary work.

Like Goethe, we made observations of form and transformation in plants, animal bones and the human skeleton. We were taught by John Wilkes, the originator of a range of elegantly sculpted surfaces called Flowforms, shaped to induce rhythmically alternating vortices in water streaming over them



A sculpture by Philip J. Kilner at Royal Brompton Hospital, London.

([www.virbelatflowforms.anth.org.uk](http://www.virbelatflowforms.anth.org.uk)). We learned to guide the movements of water and shape the swirling, fluctuating vortices, gradually gaining insight into the morphodynamics of flow. Creative engagement rarely resulted in a product that stood the test of time, but regularly led to fresh insight and discovery. I remember imagining and then shaping a variant of Wilkes' Flowform design to make a simple, asymmetric cavity that could convert continuous inflow to pulsatile outflow through a spontaneous cycle of accumulation and discharge. A pair of such forms (see above) can be seen at Royal Brompton Hospital in London.

Over the next 20 years, this discovery led me to explore the forms and dynamics of flow through heart cavities (see *Nature* **404**, 759–761; 2000), to a career in cardiovascular imaging, and to research

related to congenital heart disease and heart surgery.

But was art really significant as part of this mix, or was it just the practical approach that provided a basis for research on flow through the heart? The direct engagement with three-dimensional form and flow was relevant, but other aspects of the artistic approach were equally important.

One of these is the relative freedom to explore. Each artistic field offers its own playground for exploring the associations between phenomena. And the arts, traditionally at least, imply

aesthetic appreciation. Beauty and, perhaps, meaning tend to be apparent not through analysis but through inclusive, intuitive perception, which, like analytical and theoretical approaches, requires practice and training. Artistic exploration, whether among forms, colours or sounds, fosters a more flexible, inclusive awareness than analysis. If that kind of embracing awareness were adequately represented in scientific institutions as a counterpart to the analytical, I might feel less need to write this. But as far as I have experienced, the culture of science favours analysis at the expense of contextual awareness. The latter is at least as important, particularly when dealing with the organic, the environmental, and the human.

Philip J. Kilner is a consultant and reader in cardiovascular magnetic resonance at Royal Brompton Hospital and Imperial College, London, UK.



# Floral attraction

## An Enthusiasm for Orchids: Sex and Deception in Plant Evolution

by John Alcock

Oxford University Press: 2006. 320 pp.  
£17.99, \$29.99

### Sebebe Demissew

Botanists have devoted much time, energy and money to understanding the complexities of evolution, adaptation and speciation. In my view, orchids provide the best examples of evolutionary adaptations, because of their intricate relationships with their pollinators.

Anyone trying to learn about orchids could, however, easily be put off by the terminology and the bizarre floral characters unique to these plants. What's needed is a book that presents these complex floral characters in an understandable and attractive way. John Alcock's book *An Enthusiasm for Orchids* does this.

The book provides evidence for plant evolution at a global level, showing many captivating orchid flowers that are mainly restricted to the southwestern part of Western Australia. Spider orchids, sun orchids, hammer orchids and 'flying ducks' serve as examples of remarkable individual adaptations to their pollinators.



Designed to deceive: an Australian bee orchid, *Caladenia discoidea*.

Alcock explains why this part of Australia has such diversity and discusses the speciation.

Since the publication of Charles Darwin's *The Origin of Species*, many authors have provided examples in support of natural selection, and others have tried to refute it. This book certainly belongs in the former category, providing vivid examples of adaptation shaped by natural selection promoting reproductive success.

Field biologists face many problems on field trips, such as difficulty in finding the right location and the species they are looking for, and not being sure whether the species they find is new to science or not. Few researchers would

transmit this information to others, but Alcock gives a full account of his experience that will be useful for anyone wishing to explore the south-west corner of Australia. I visited Western Australia in the 1980s and was fascinated by the banksias and sundews, for example, that Alcock also mentions. Although I am currently more involved with petaloid monocots, such as aloes, with my interest in orchids this book has inspired me to revisit southwestern Australia when the opportunity arises.

The book is easy to read and has many beautiful illustrations. It shows adaptations found in different groups of plants: orchid flowers lure their pollinators by various means, sundews have folding leaves to capture insects as prey, and plants also rely on winged seeds or the explosion of dried capsules for dispersal.

For an overview of the concepts of adaptation and maladaptation, a brief history of evolution in general, and a good look at the hotspots, biodiversity and conservation of orchids in southwestern Australia in particular, this is the book to read.

Sebebe Demissew is keeper of the Ethiopian National Herbarium and heads the Ethiopian Flora Project, Addis Ababa University, P.O. Box 3434, Addis Ababa, Ethiopia.

# The maths mentor

## Letters to a Young Mathematician

by Ian Stewart

Basic Books: 2006. 224 pp. \$22.95, £13.99

### Rhian Parker

Mathematics seems to be something that you either get or you don't. I was lucky: I've always found the concepts behind maths easy to understand, and in some way natural. A degree in maths was always what I was going to do, a master's followed, and the possibility of staying in academia is something I've been considering since I left school. There are already a number of 'popular' maths books (many by Ian Stewart), which gave me a good idea of what the transition from school to university maths would be like. But I could do with more information on the more personal aspects of how one goes about doing maths, and the possibilities for being a professional mathematician, even as I approach the end of my master's.

In *Letters to a Young Mathematician*, Stewart writes to a fictional girl, Meg, following her career in mathematics from high school to a tenured position. Each letter covers a different topic, answering concerns that Stewart feels young mathematicians may have.

He starts on familiar ground with the differences between school and university mathe-

matics, and moves on to what to expect from a maths degree. These are mainly discussions on proof — what it is, why we feel the need to do it, and how Stewart likes to think of it. As Meg progresses, the letters move on to what professional mathematicians do and how they do it. A recurring theme is what maths actually is, with letters at times verging on philosophy.

As a young female mathematician, I found the book somewhat disappointing. It offers good advice in some areas, such as how to work at maths at university level. But there is a general lack of information — Stewart tends to waffle slightly. For instance, the letter about the career ladder is written in the context of finding a good PhD supervisor. Nowhere does he describe the actual professional ladder and where you can expect a career in mathematics to take you — information that would be welcome to a student deciding on a degree subject, and to an undergraduate deciding between an academic career and moving out. There is also too much jargon: Meg gets a postdoc position, then an assistant professorship, and ends up with tenure. I wouldn't have understood any of these terms as a 17-year-old.

The book also suffers from being written entirely for a US audience, from the spelling to the description of how your career pro-

gresses. This is a real shame, as the differences are at times enough to make the book all but useless to students outside the United States.

The letter in which Stewart tells Meg how to teach undergraduates should be compulsory reading for all lecturers and tutors. The advice seems to get better the further Meg gets. There is an amusing chapter on learning from others' mistakes, and his description of the community of mathematicians is true and appealing.

But the book's major problem is the timescale. Meg ages 15 or 20 years; the reader a few hours. The first few chapters are good for school pupils and as an introduction to 'higher level' mathematics. The next few can be read when you're getting to the end of your degree, and then again when you gain your PhD. But at any one time, the book has too little to offer, and the rest of it is either below you or not written in a general enough way to interest you.

And what saddened me was that *Letters to a Young Mathematician* did not leave me with a sense of wonder and beauty, or even pride, in mathematics, as other books have. For that I would recommend Paul Hoffman's *The Man Who Loved Only Numbers* (Fourth Estate, 1998), Robert Kanigel's *The Man Who Knew Infinity* (Scribner, 1991) and most of Stewart's other books, especially *The Magical Maze* (Weidenfeld & Nicolson, 1997).

Rhian Parker is fourth-year mathematics student at The Queen's College, University of Oxford, Oxford OX1 4AW, UK.

## GENE REGULATION

# A finger on the mark

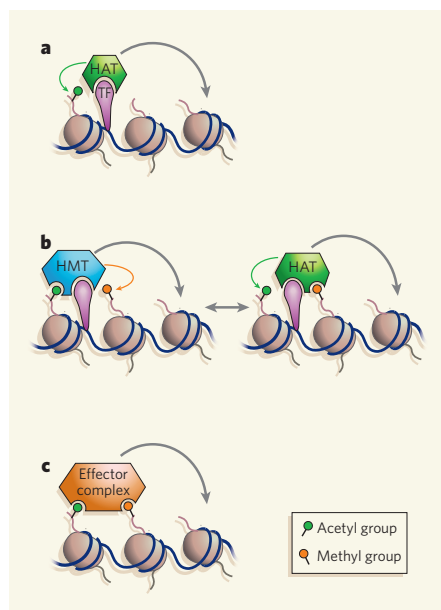
Peter B. Becker

**Chromatin, the protein wrapping of the genome, harbours information about how the genes it contains are to be regulated. Are we any closer to deciphering these encoded instructions?**

From our everyday shopping, we are familiar with the fact that the wrapping of our groceries is decorated with information about the purchased goods. Some of this information, such as the bar code, can be accessed only with special decoding equipment. Likewise, the protein wrapping that stows and organizes DNA in the cell nucleus may be marked chemically as a way of indexing the underlying genes. Four papers in this issue<sup>1–4</sup> illustrate how a methyl mark is 'read' by interacting proteins, which convert the signal into function.

The packaging, organization and indexing of genomes in the nucleus is achieved through their association with a large number of proteins to form a complex structure called chromatin. First and foremost, short DNA segments are wound around spools made of histone proteins to form nucleosomes. The chromatin fibre consists of long arrays of such nucleosomes, which coil into higher-order structures that are not entirely understood. The folding of the chromatin fibre — conceivably through its compactness and rigidity — can be tuned to either permit or prevent the use of the genetic information. This is largely determined by molecular interactions of the histones, which, in addition to the globular parts that comprise the DNA spool, have long 'tail' domains that reach out to contact other proteins contributing to structure or function. These interactions may be disturbed or promoted by chemical modifications of the histone tail — the most common of which is the addition or removal of phosphate, methyl or acetyl groups. According to prevailing ideas, the status of a cell may be signalled to chromatin through the placement of such chemical tags. The marked chromatin is then recognized by 'effector' proteins that regulate how the underlying genetic information is used<sup>5</sup>.

The regulators of chromatin structure recognize histone marks through dedicated protein domains. Prominent examples are the so-called bromo- and chromodomains, which are able to bind, respectively, to histone tails that carry acetyl or methyl marks at certain amino-acid residues<sup>5,6</sup>. For example, a triple methylation mark on lysine 4 of histone H3 (designated



**Figure 1 | Recruitment of chromatin regulators by nucleosomes.** These hypothetical scenarios show nucleosomes as brown spheres around which the DNA (blue) is wound. **a**, A DNA-bound transcription factor protein (TF) may be the initial contact for chromatin regulators, in this case a histone acetyltransferase (HAT). The TF binds to specific DNA sequences so that the HAT complex is recruited to particular regions of the DNA. The HAT enzyme adds an acetyl group to the nearby histone tails and may have additional functions in chromatin regulation (black arrow). **b**, Regulators may also be targeted to particular DNA regions by a combinatorial interaction with a transcription factor and an acetylation mark. Here a histone methyltransferase complex (HMT) adds a methyl group to nearby histone tails, among other possible effects (black arrow). The alternating recruitment of HAT and HMT reinforces the complex histone modification pattern. **c**, A combination of histone modification marks, in the absence of a transcription factor, may suffice to stably recruit a regulatory protein complex to a regulatory site.

H3K4me3) is a hallmark of all active genes.

Last year, several groups provided the first insight into how this triple methyl mark can contribute to gene activation<sup>7–9</sup>, by showing that the nucleosome remodelling factor CHD1 binds preferentially through its chromo-domain to nucleosomes carrying the mark. Nucleosome remodelling factors are enzymes that can disrupt histone–DNA interactions such that the nucleosomes release crucial regulatory sequences of DNA. Guiding such a factor to the nucleosomes that organize gene control regions will lead to local opening of the chromatin and hence possibly to activation. CHD1 may also associate with protein complexes containing an enzyme that tags histones with an acetyl group (histone acetyltransferase), and may recruit the complexes to the H3K4 methyl mark<sup>8</sup> (Fig. 1b). Because histone acetylation leads to chromatin unfolding, an H3K4 methylation signal can coordinate 'downstream' chromatin modifications on a path towards opening a repressive chromatin structure. Interestingly, 'double tudor' domains, which belong to the same super-family as chromodomains, can also target chromatin

regulators to distinct methylation marks on histones<sup>10</sup>.

The work presented in this issue has uncovered a novel principle by which an H3K4 methyl mark can be recognized and translated into gene-expression patterns. Wysocka *et al.* (page 86)<sup>1</sup> observed that a nucleosome remodelling factor called NURF can be recruited to H3K4 trimethylated chromatin through a selective interaction with NURF's largest subunit, BPTF. The recruitment of NURF through the H3K4me3 mark is vital for the faithful activation of developmental genes, and interference with either the establishment of the methylation mark or its recognition by NURF leads to defects in frog embryonic development<sup>1</sup>. But BPTF does not contain a chromodomain, so how does it interact with the chromatin?

Detailed molecular analysis showed that BPTF recognizes the methyl mark through a structure called the PHD finger, a well-known protein structural fold coordinated by two zinc atoms. PHD fingers are abundant in chromatin regulators<sup>11</sup>, but their function had not been identified. It seems, however, that an



affinity for methylated histone tails is a general feature of PHD fingers, because Shi *et al.* (page 96)<sup>2</sup> observe that the PHD fingers of the tumour-suppressor protein ING2 (for 'inhibitor of growth') and those of related proteins interact preferentially with methylated lysine 4 of histone H3. However, whereas the recruitment of NURF or CHD1 leads to gene activation, when ING2 is recruited it brings with it a histone deacetylase (HDAC) that removes nearby acetylation marks, triggering a closed chromatin structure in which the genes are repressed. ING2 inhibits cell proliferation when genomic DNA is damaged, so Shi *et al.* examined the effects of DNA damage on the interaction of ING2 with chromatin. Experimental DNA damage led to binding of the ING2–HDAC complex to the regulatory region of the gene encoding cyclin D1, a major inducer of cell proliferation, and the expression of this gene was repressed.

BPTF and ING2 interact preferentially with nucleosomes methylated at lysine 4 *in vitro* and will not bind to nucleosomes methylated at lysine 9. The molecular basis for this selectivity is clearly illustrated by the structures of the PHD fingers of these proteins in complex with the methylated peptides. These structures have been determined by Li *et al.* (page 91)<sup>3</sup> and Peña *et al.* (page 100)<sup>4</sup>. The PHD fingers form surface pockets that precisely accommodate a short region of the H3 tail, including methylated lysine 4, but not an analogous segment surrounding methylated lysine 9. A common principle of all methyl recognition events is the presence of two to four aromatic amino acids that form a hydrophobic environment<sup>3,4,10</sup>. Variations in PHD finger structure allow other methylated peptides to be accommodated. For example, the PHD finger of the nucleosome remodeller Mi2 interacts preferentially with H3 tails methylated at lysine 36 (ref. 2).

Those who had hoped for a clear-cut set of rules governing the recognition and interpretation of methyl marks — a 'histone code' — will be disappointed by the embarrassment of riches uncovered by the latest reports. Given the multitude of potential interactors that bring about diverse functional consequences, and the relative weakness of histone-tail interactions, it is clear that histone methylation alone cannot specify the selectivity of nucleosome interactions observed *in vivo*. Additional specificity could be provided by combinations of histone modifications that may be recognized by corresponding combinations of interacting domains. For example, the PHD finger of BPTF may cooperate with a neighbouring bromodomain in docking with a combined methyl–acetyl mark<sup>3</sup> (Fig. 1c). On the other hand, alternative recruitment strategies involving sequence-specific DNA-binding proteins are known for all the regulators mentioned here. So it is likely that histone methylation marks are not the primary targeting determinants, but rather contribute to the stabilization or concentration of factors once

they have been targeted to a certain chromatin neighbourhood (Fig. 1a,b).

We have only just begun to unravel the pathways towards chromatin opening that integrate the various writers and interpreters of histone modification marks<sup>10,12</sup>. Decoding the information hidden in the chromatin wrapping promises to reveal the instructions for how the packaged genetic information is to be used.

Peter B. Becker is at the Adolf-Butenandt-Institut, Molekularbiologie, Schillerstrasse 44, 80336 München, Germany.

e-mail: pbecker@med.uni-muenchen.de

1. Wysocka, J. *et al.* *Nature* **442**, 86–90 (2006).
2. Shi, X. *et al.* *Nature* **442**, 96–99 (2006).
3. Li, H. *et al.* *Nature* **442**, 91–95 (2006).
4. Peña, P. V. *et al.* *Nature* **442**, 100–103 (2006).
5. Schreiber, S. L. & Bernstein, B. E. *Cell* **111**, 771–778 (2002).
6. Turner, B. M. *Cell* **111**, 285–291 (2002).
7. Sims, R. J. III *et al.* *J. Biol. Chem.* **280**, 41789–41792 (2005).
8. Pray-Grant, M. G., Daniel, J. A., Schieltz, D., Yates, J. R. III & Grant, P. A. *Nature* **433**, 434–438 (2005).
9. Flanagan, J. F. *et al.* *Nature* **438**, 1181–1185 (2005).
10. Huang, Y., Fang, J., Bedford, M. T., Zhang, Y. & Xu, R. M. *Science* **312**, 748–751 (2006).
11. Bienz, M. *Trends Biochem. Sci.* **31**, 35–40 (2006).
12. Wysocka, J. *et al.* *Cell* **121**, 859–872 (2005).

## CONDENSED-MATTER PHYSICS

# Superfluidity in the picture

J. E. Thomas

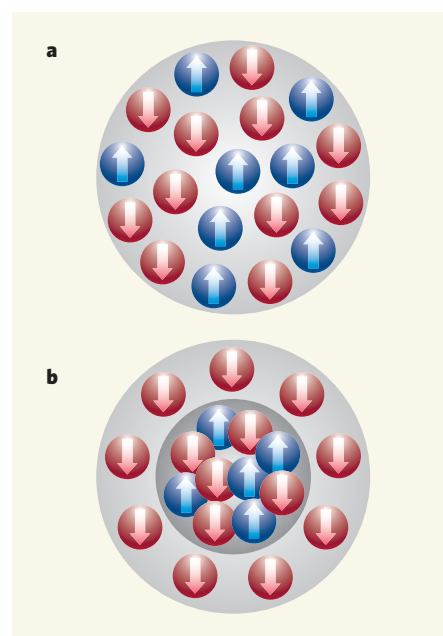
**Strongly interacting atomic Fermi gases — useful models for many other exotic forms of matter — enter a superfluid state at low temperatures. The first direct observation of that transition has been made.**

Fermions are the building-blocks of the Universe. Strongly interacting gases of these particles make up atomic nuclei, the matter in neutron stars and the quark–gluon plasma of the Big Bang. These systems differ greatly in size and energy but have a lot in common; knowledge of one system therefore illuminates our understanding of the others.

On page 54 of this issue, Zwierlein *et al.*<sup>1</sup> present direct observations of a superfluid phase in one such system — a gas of strongly interacting Fermi atoms — and also determine the temperature of the gas directly. Such atomic gases have been the subject of intensive study<sup>2–5</sup>, as they are particularly easy to create and control in a laboratory environment. Although it is accepted that these atoms have been cooled down to an intriguing type of superfluid state, a direct observation of this transition, similar to the formation of ice from freezing water, has been elusive.

The classification of fundamental particles into fermions and bosons depends on their internal rotation, or spin. Spin is a quantized property; that is, it can only take on certain discrete values. If a particle's spin is a half-integer multiple of Planck's constant  $h$  divided by  $2\pi$ , it is a fermion. The familiar atomic building-blocks — the electron, neutron and proton — are all fermions. Spin can point 'up' or 'down', and fermions obey the Pauli exclusion principle, which permits at most one spin-up and one spin-down particle in each quantum energy level. This explains the way in which electrons are observed to fill up the energy levels in atoms, two by two.

Bosons, on the other hand — particles of whole-integer spin — experience no such restriction. Composite particles (atoms, for example) that are built out of an even number



**Figure 1 | Changing shape.** **a**, A homogeneous state of an unequal number of spin-up and spin-down atoms, such as that prepared by Zwierlein *et al.*<sup>1</sup>. **b**, As the gas temperature is lowered past a critical value, a phase transition occurs. A superfluid state forms in the centre of the gas with spin-up and spin-down atoms paired and equal in number, and the excess unpaired atoms move to outside the central core.

of fermions are bosons, while those containing an odd number of fermions are fermions. So whereas a gas of atomic fermions stacks up in the energy levels in a ladder-like fashion, with at most one spin-up and one spin-down atom on each rung, all the atoms in an ultracold gas of bosons can condense into the lowest quantum energy state: the superfluid state of matter

known as a Bose–Einstein condensate. For fermions, this is generally not possible, but this situation changes dramatically when spin-up and spin-down particles attract each other. Then, at sufficiently low temperature, fermions of opposite spin can pair to form composite bosonic objects, and a superfluid that flows without friction is also produced in this case.

In metals, the pairing of spin-up and spin-down electrons produces a superconducting superfluid, in which electric current can flow without resistance. If the strength of the pairing between the electrons is weak, and the pairing energy is a small fraction of the electrons' total energy, this superconducting transition occurs at very low temperatures (below the boiling point of helium at 4.22 K), and thus is useful only in a laboratory environment. A strong pairing, by contrast, can produce a transition at high temperature. Super-high-temperature superconductors that would operate at temperatures well above room temperature would enable lossless power transmission, magnetic levitation and more efficient communication. The interest in developing such materials is understandably great.

Remarkably, certain atomic Fermi gases at temperatures of a fraction of a millionth of a degree above absolute zero provide a model system for testing the theory of such high-temperature superconductors, because the pairing energy of their atoms is a large fraction of the fluid's total energy — in fact larger than that in any existing superconductor. Such gases have a feature known as a Feshbach resonance, near which the strength of the interactions between spin-up and spin-down atoms can be tuned over a wide range simply by applying a magnetic field. At resonance, the attractive interaction between atoms is so strong that the gas flows like a high-temperature superfluid, modelling the movement of electrons in a superconductor that would work even at temperatures of thousands of degrees. Many experiments, including measurements of pair condensation<sup>6,7</sup>, collective oscillation modes<sup>8,9</sup>, pairing energy<sup>10</sup>, heat capacity<sup>11</sup> and, most recently, vortices that signal superfluid flow<sup>12</sup> have established that the transition to a superfluid phase occurs. But until now, the direct signature of the transition that would be given by a change in shape of the atomic cloud had not been observed.

Zwierlein and colleagues' proof<sup>1</sup> of that change comes from a mixture of fermionic lithium-6 atoms in which the number of spin-up and spin-down atoms is unequal<sup>13,14</sup>. At temperatures above a critical value, the trapped mixture remains homogeneous, with a uniform distribution of both spin-up and spin-down particles across the trapped mixture (Fig. 1a). As the temperature is lowered, however, the gas cloud suddenly makes a transition to an energetically more favourable shape. A higher-density central 'bump' forms that contains nominally equal numbers of paired spin-up and spin-down atoms, while

the excess of the majority spin component moves to the outside (Fig. 1b).

This abrupt change is interpreted as a phase transition, analogous to the formation of ice in water when cooled to below its freezing point. When the gas is rotated, vortices form in the central region. Such vortices require perfect hydrodynamic flow, so their appearance is a sure sign that this region is a superfluid.

The majority spin component in the outside region brings an added benefit: it functions as a thermometer for the gas. When the trap initially confining the atoms is turned off, this pure component expands ballistically. That enables precise measurement of the velocity distribution of the atoms, which in turn is a measure of the temperature of the gas. As most thermodynamic quantities depend on temperature, this provides a valuable experimental check on theoretical models.

With precise control of spin populations and precision thermometry, the experiments open up new territory for modelling some of the most fundamental processes in the Universe, including superfluidity in neutron stars, hydrodynamics in a quark–gluon plasma

and string-theory predictions of minimum viscosity in any generalized system. ■

J. E. Thomas is in the Department of Physics, Duke University, Durham, North Carolina 27708, USA.  
e-mail: jet@phy.duke.edu

1. Zwierlein, M. W., Schunck, C. H., Schirotzek, A. & Ketterle, W. *Nature* **442**, 54–58 (2006).
2. O'Hara, K. M., Hemmer, S. L., Gehm, M. E., Granade, S. R. & Thomas, J. E. *Science* **298**, 2179–2182 (2002).
3. Thomas, J. E. & Gehm, M. E. *Am. Sci.* **92**, 238–245 (2004).
4. Kinast, J., Turlapov, A. & Thomas, J. E. *Opt. Phot. News* **16**, 21 (2005).
5. Chevy, F. & Salomon, C. *Phys. World* **18**, 43–47 (2005).
6. Regal, C. A., Greiner, M. & Jin, D. S. *Phys. Rev. Lett.* **92**, 040403 (2004).
7. Zwierlein, M. W. *et al. Phys. Rev. Lett.* **92**, 120403 (2004).
8. Kinast, J., Hemmer, S. L., Gehm, M. E., Turlapov, A. & Thomas, J. E. *Phys. Rev. Lett.* **92**, 150402 (2004).
9. Bartenstein, M. *et al. Phys. Rev. Lett.* **92**, 203201 (2004).
10. Chin, C. *et al. Science* **305**, 1128–1130 (2004).
11. Kinast, J. *et al. Science* **307**, 1296–1299 (2005).
12. Zwierlein, M. W., Abo-Shaeer, J. R., Schirotzek, A., Schunck, C. H. & Ketterle, W. *Nature* **435**, 1047–1051 (2005).
13. Partridge, G. B., Li, W., Kamar, R. I., Liao, Y. & Hulet, R. G. *Science* **311**, 503–505 (2006).
14. Zwierlein, M. W., Schunck, C. H., Stan, C. A., Raupach, S. M. F. & Ketterle, W. *Science* **311**, 492–496 (2006).

## VIROLOGY

# Micro mystery solution

Bill Sugden

**A particular talent of herpes simplex virus-1 is that it can lurk unseen in the cells of an infected person for long periods. It turns out that the virus achieves this feat through the agency of a microRNA.**

Herpes simplex virus-1 (HSV-1) infects most of us but then stays hidden for most of our lives. When it does re-emerge, the virus usually causes cold sores; its close cousin, HSV-2, can likewise hide, then re-emerge to cause distressing and painful genital sores. How these viruses stay hidden in the face of a vigorous immune response is a mystery that has yielded only fitfully to investigation. Nigel Fraser and his colleagues (page 82 of this issue)<sup>1</sup> have now identified a small HSV-1 gene that helps infected neurons to survive, and thereby enables the virus to remain hidden in them.

The genome of HSV-1 contains about 100 genes. The virus expresses most of them in cells in oral sores. Infection leads to amplification of its DNA, formation of viral particles that encapsulate the new viral DNA (Fig. 1, overleaf), death of the host cells, and release of progeny viral particles. This 'productive phase' of infection is common in viruses. What is unusual about herpes viruses is that they also have a latent phase, during which they reside in some cells of an infected organism without supporting a productive infection.

HSV-1 infects people. But learning how this pathogen co-habits with us has required the

use of animal hosts as models for experimental analysis. Two decades ago, mice that survived an initial infection with HSV-1 were found to harbour HSV-1 DNA in their trigeminal ganglia, as we do<sup>2</sup>. The trigeminal ganglia are the clusters of neurons that innervate our face, with some neurons terminating near the lips. It turned out that these infected neurons predominantly or solely express just one transcript of the viral genome — a very unexpected result<sup>3</sup>. That transcript was also identified in latently infected neurons in humans, and was called the latency-associated transcript, or LAT. As it was the sole viral participant in latency, LAT was assumed to have a pivotal role in the phenomenon.

The ensuing analyses yielded perplexing findings, however. The LAT gene is initially transcribed to yield an RNA that is 8.5 kilobases long but lacks the hallmarks of being a precursor to a messenger RNA<sup>4</sup>. It is processed, or spliced, as are precursors to mRNAs, but surprisingly a 2-kilobase stretch derived from the splicing is long-lived and tends to remain in the host cell's nucleus<sup>4</sup>. No protein encoded by LAT has been consistently detected *in vivo*.





## 50 YEARS AGO

"Atomic time and the definition of the second" — The following tentative proposals are put forward in an attempt...to make the atomic standard immediately available but to preserve a single unit of time closely linked with, although not defined by, that given by astronomical observations.

(1) The provisional unit of time should be the time of 9 192 631 830 cycles of the caesium  $Fm(4,0) \leftrightarrow (3,0)$  resonance at zero field, and named 1 provisional second...

In making these proposals, the understandable reluctance to break with a practice as old as civilization is fully appreciated; but a consideration of the precision of measurement alone shows that the astronomical unit of time is no longer sufficiently accurate for modern measurements of frequency and time interval.

From *Nature* 7 July 1956.

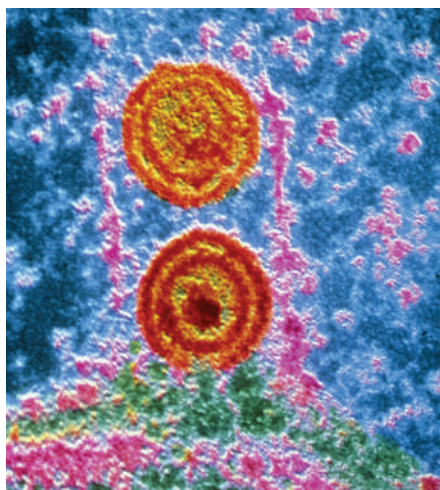
## 100 YEARS AGO

"The olfactory sense in *Apteryx*"

— About a year ago I stated in your columns that I was trying to have some experiments carried out with the object of ascertaining whether the olfactory sense of the kiwi [and roa] is perceptibly developed, as one would suppose it to be from...the great relative size of the olfactory lobes of the brain and the great size of the olfactory capsule as seen in the skull. I wrote to the curators of Little Barrier and Resolution Islands asking each of them to try certain experiments for me... Mr. Henry placed a number of earthworms at the bottom of shallow buckets and covered them with four inches of earth. When such a bucket was placed on the ground the roa got quite excited in its hunt through the earth, probing to the bottom for the worms. It must be borne in mind that the roa (and kiwi) is practically blind during the day...while Mr. Henry states that it makes such a "sniffing noise" that it would be unable to hear a worm... "when I put down a bucket of earth without worms in it, the bird would not even try it".

From *Nature* 5 July 1906.

INST. PASTEUR/UNITE VIRUS ONCOGENES/SPL



**Figure 1 | Active herpes simplex virus.** This false-colour electron micrograph shows virus particles (orange) migrating from the host cell's nucleus. ( $\times \sim 190,000$ .)

Moreover, it was found that when mice were exposed to a mutant of HSV-1 lacking *LAT*, the virus infected trigeminal ganglia but escaped from latency only inefficiently<sup>5</sup>. Subsequently, a sensitive quantitative assay showed that HSV-1 lacking *LAT* was inefficient in infecting neurons latently<sup>6</sup>. Thus, *LAT* seemed to be important in both HSV-1 entry into and escape from latent infections.

Some of the mystery was solved when uninfected and infected neurons could be counted accurately. About a quarter of the neurons of a trigeminal ganglion are killed following infection with HSV-1, but more — half — are killed when *LAT* is absent<sup>7</sup>. Similarly, without *LAT*, proportionately fewer neurons survive to be latently infected<sup>7</sup>. Thus, *LAT* might not contribute to HSV-1 escape from latency but it does help neurons survive infection. It promotes neuronal survival by inhibiting apoptosis, or programmed cell death, the usual process by which an infected cell self-destructs in order to prevent production of viral progeny. For example, the introduction of *LAT* in cultured cells inhibits apoptosis induced by several agents<sup>8</sup>; and substitution of *LAT* in HSV-1 with a gene known to inhibit apoptosis recapitulates *LAT*'s role *in vivo*<sup>9</sup>.

What has remained a mystery is how *LAT* promotes neuronal survival. This is where Fraser and colleagues<sup>1</sup> come in. Examination of the sequence of the *LAT* gene revealed a stretch of DNA — an inverted repeat sequence — that could be recognized by the host cell as a precursor to a microRNA (miRNA). MicroRNAs are small RNAs, first identified in plants, that regulate gene expression after the gene has been transcribed into mRNA; they do this either by degrading the mRNA or by inhibiting its translation into protein. Fraser and co-workers detected the predicted miRNA, miR-LAT, in cells infected with HSV-1 but not in those infected with a virus defective in its *LAT* gene. This viral miRNA evidently inhibits apoptosis induced in cells in culture — but how?

Finding the answer was no easy task. MicroRNAs share only imperfect complementarity with sequences in the control regions — the 3' untranslated regions — of the mRNAs that they regulate. So searching for their target mRNAs is difficult, because the limited sequence that they have in common means there are many candidates. Fraser and colleagues succeeded in this quest by considering genes that had appropriate functions and for which the shared sequences are evolutionarily conserved, a feature common to sequences targeted by miRNAs. They found that miR-LAT promotes the degradation of mRNAs encoding two proteins — transforming growth factor- $\beta$  (TGF- $\beta$ , which can both inhibit cell proliferation and induce cell death) and SMAD3 (a mediator of the signalling induced by TGF- $\beta$ ). In other words, it looks as if miR-LAT short-circuits the ability of an infected cell to self-destruct.

Overall, we now have the following picture of HSV-1 biology. The virus infects us, wends its way to trigeminal neurons and supports productive infection there. That infection elicits a vigorous cellular immune response, which kills some of the infected cells. In others, however, a latent infection becomes established<sup>10</sup>. Occasionally, HSV-1 escapes from latency to grow productively, killing its host neuron. These re-emergences are both controlled by and continue to attract the nearby immune cells.

The HSV-1 *LAT* gene means that latently infected neurons are more likely to survive in this environment. These neurons express no viral proteins to be recognized by the host's immune system — instead they express miR-LAT, which means that the neuron is more likely to survive, and so provide a hiding place for the virus.

Evolutionary relatives of mammalian miRNAs are major regulators of gene expression in species that have been studied closely. For example, in plants and worms they mediate systemic gene silencing or control specific steps in development. The similarities in synthesis and function between the miRNAs of plants, invertebrates and mammals make it certain that they have roles in us that we are only beginning to glimpse.

Bill Sugden is in the McArdle Laboratory for Cancer Research, University of Wisconsin, Madison, Wisconsin 53706, USA.  
e-mail: sugden@oncology.wisc.edu

- Gupta, A., Gartner, J. J., Sethupathy, P., Hatziageorgiou, A. G. & Fraser, N. W. *Nature* **442**, 82–85 (2006).
- Rock, D. L. & Fraser, N. W. *Nature* **302**, 523–525 (1983).
- Stevens, J. G., Wagner, E. K., Devi-Rao, G. B., Cook, M. L. & Feldman, L. T. *Science* **235**, 1056–1059 (1987).
- Jones, C. *Clin. Microbiol. Rev.* **16**, 79–95 (2003).
- Leib, D. A. et al. *J. Virol.* **63**, 2893–2900 (1989).
- Thompson, L. & Sawtell, N. M. *J. Virol.* **71**, 5432–5440 (1997).
- Thompson, L. & Sawtell, N. M. *J. Virol.* **75**, 6660–6675 (2001).
- Perng, G. C. et al. *Science* **287**, 1500–1503 (2000).
- Jin, L. et al. *J. Virol.* **79**, 12286–12295 (2005).
- Theil, D. et al. *Am. J. Pathol.* **203**, 2179–2184 (2003).

## MICROSCOPY

## X-ray nanovision

Eric D. Isaacs

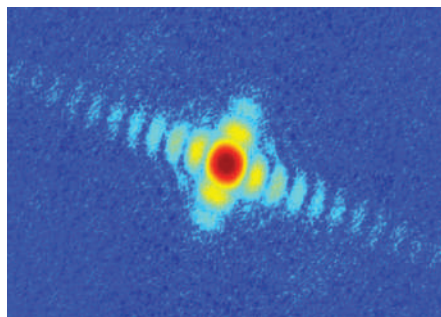
**Startling three-dimensional images of nanoparticles have been obtained with an X-ray microscope, showing crystal deformation in unprecedented detail. The trick is not to focus the X-rays, but to diffract them.**

Imagine an imaging device with resolution more than 1,000 times better than an optical microscope and capable of seeing deep inside objects. Such is the promise of an X-ray microscope, which has been the dream of scientists since X-rays were discovered over a century ago. On page 63 of this issue<sup>1</sup>, Ian Robinson and colleagues present a significant step towards realizing this dream, using the technique of coherent X-ray diffraction imaging. Their results demonstrate the power of X-ray microscopy to create three-dimensional images at the nanoscale, and indicate a clear path towards the ultimate X-ray microscope with atomic-scale resolution.

There is an acute need in nanotechnology for images with atomic-scale resolution. Because nanoparticles consist of relatively few atoms, their mechanical, electrical and thermodynamic behaviours are mainly determined by the particles' surfaces and interfaces with other materials, by defects such as those caused by atomic vacancies or non-native atoms, and by deformation of the atomic crystal lattice, known as strain. For example, nanoparticles of gold, which consist of just thousands of atoms, are excellent catalysts for certain chemical reactions, even though bulk gold is essentially inert<sup>2</sup>. Semiconductor nanocrystals, known as quantum dots, have electronic and optical behaviour that can be tuned by binding a single molecule to their surface<sup>3</sup>. To develop materials with useful properties, we must first understand how these properties relate to atomic structure and surface chemistry.

Electron microscopes are ubiquitous in materials science, but X-ray microscopes are emerging as an attractive alternative, for two reasons. First, the penetrating power of X-rays into materials allows us to quantitatively determine the full three-dimensional structure of nanoparticles, usually non-destructively. High-resolution electron microscopy often requires samples to be thinned to several atomic layers. This is technically challenging for many materials and can alter the properties being studied. Embedded nanostructures can best be studied with X-rays.

Second, bright synchrotron sources of 'hard' X-rays — those with a wavelength shorter than 0.1 nm — have become readily available. This has enabled the development of X-ray optics with improved resolving power. So far, microscopes based on Fresnel zone plates (which focus light by diffraction) or mirror optics



**Figure 1 | Coherent X-ray diffraction.** The pattern created by the diffraction of coherent X-rays from a 160-nm cube of silver<sup>10,11</sup>. This may be mathematically 'inverted' to produce an image of the cube.

have produced the highest resolution (about 40-nm) hard X-ray images. But the optics used by these methods are technologically challenging to make<sup>4,5</sup>. Coherent X-ray imaging, which uses in-phase X-rays, offers a 'lens-less' alternative.

When coherent X-rays irradiate an object, the scattered radiation forms a pattern. In contrast to the scattering produced with incoherent X-rays, the coherent pattern contains not only information about the amplitude of the scattered radiation, but also information about the phase of that radiation, provided that the pattern is sufficiently 'oversampled'<sup>6</sup>. Oversampling is a well-established technique used to ensure proper retrieval of phase information from the scattered radiation. Crucially, the phase information allows the pattern to be mathematically 'inverted' to recover an image of the charge-density distribution of the object. One significant feature of coherent-diffraction imaging is its high sensitivity to crystallographic strain — that is, to deviations of atoms from their nominal bulk positions.

Coherent X-ray imaging was proposed in the early 1980s as a means to visualize the electron-density distribution in non-crystalline materials<sup>7</sup>, and was first demonstrated in 1999 (ref. 8). Two years later, the technique was used by Robinson *et al.*<sup>9</sup> on a crystalline system. Figure 1 shows an example of the coherent diffraction pattern from a 160-nm silver cube, observed in previous work by Robinson and colleagues<sup>10,11</sup>. The four symmetric streaks are signatures of the cubic facets. These streaks are produced by sharply terminated surfaces and could be formed with incoherent X-rays. Coherence is, however, required to produce the periodic modulation of the streaks.

The modulation is an interference pattern created by X-rays coherently scattered from opposing sides of the cubes, much like the interference pattern in the classic two-slit experiment of Thomas Young. This pattern was inverted to create an image of the cube<sup>10,11</sup>.

Robinson and colleagues<sup>1</sup> have now applied their technique to more complicated structures. They collected the coherent diffraction pattern from 750-nm hemispherical lead nanoparticles that had small faceted faces. The pattern was then inverted by applying computational techniques developed by the authors, to obtain a full three-dimensional image of the electron-density distribution in the particles.

A truly remarkable aspect of this image is the clear articulation of a deformation of the lattice, known as a strain field (see Figs 3 and 4 of ref. 1). When contour maps of the magnitude of the strain are plotted, it becomes apparent that the strain field originates from a point just outside the lead particle, in the region of contact with the glass substrate upon which the particle was grown. This strongly supports the authors' claim that the strain field is caused by interfacial contact forces between the particle and the substrate. These forces can be associated with a single defect that may have been the nucleation point for the growth of the nanocrystal.

By creating a lens-less image of the strain fields in a lead nanoparticle with 40-nm spatial resolution, Robinson and co-workers have taken a substantial step forward in realizing the potential of X-ray microscopes. One can envisage even greater uses, for instance imaging the strain field near a single atomic defect inside a semiconductor quantum dot. Such applications will require significant improvements in detectors and a much higher flux of coherent X-rays, which will become available with future generations of hard X-ray sources, such as the X-ray Free Electron Laser. With these advances, the dream of a microscope capable of imaging the position and type of every atom in a nanocrystal or macromolecular assembly could become a reality. ■

Eric D. Isaacs is in the Argonne National Laboratory, Center for Nanoscale Materials, Argonne, Illinois 60439, USA, and the James Franck Institute, University of Chicago.  
e-mail: isaacs@anl.gov

1. Pfeifer, M. A., Williams, G. J., Vartanyants, I. A., Harder, R. & Robinson, I. K. *Nature* **442**, 63–66 (2006).
2. Haruta, M. in *Encyclopedia of Nanoscience and Nanotechnology* Vol. 1 (ed. Nalwa, H. S.) 655–664 (Am. Sci. Publ., Stevenson Ranch, CA, 2004).
3. Rajh T. *et al.* *J. Phys. Chem. B* **106**, 10543–10552 (2002).
4. Evans, P. & Isaacs, E. D. *J. Phys. D* **39**, R1–R19 (2006).
5. Jacobsen, C. & Kirz, J. *Nature Struct. Biol.* **5**, 650–653 (1998).
6. Shen, Q., Bazarov, I. & Thibault, P. *J. Synch. Rad.* **11**, 432–438 (2004).
7. Sayre, D. *Lect. Not. Phys.* **112**, 229–235 (1980).
8. Miao, J., Charalambous, P., Kirz, J. & Sayre, D. *Nature* **400**, 342–344 (1999).
9. Robinson, I. K. *et al.* *Phys. Rev. Lett.* **87**, 195505 (2001).
10. Robinson, I. K. *et al.* *Opt. Express* **11**, 2329–2334 (2003).
11. van der Veen, F. & Pfeiffer, F. *J. Phys. Condens. Matter* **16**, 5003–5030 (2004).

I. K. ROBINSON *ET AL.*



## BRIEF COMMUNICATIONS

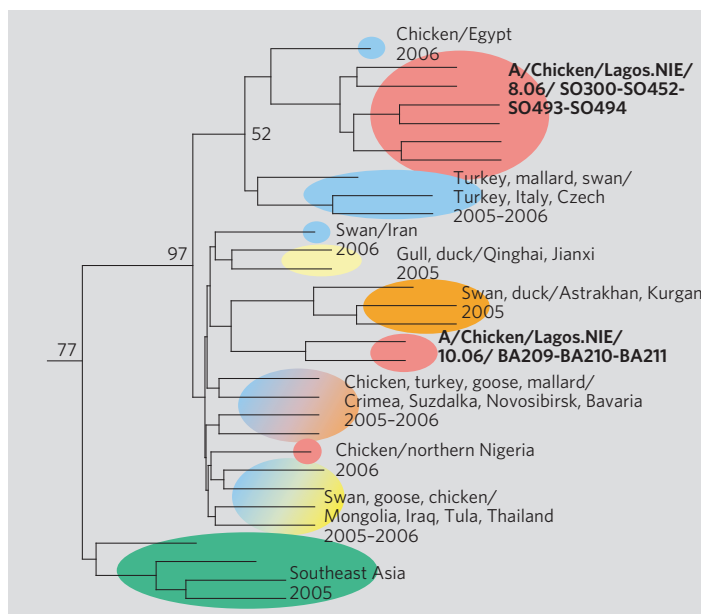
## Multiple introductions of H5N1 in Nigeria

Phylogenetic analysis reveals that this deadly virus first arrived in Africa from different sources.

As the avian influenza virus H5N1 swept from Asia across Russia to Europe, Nigeria was the first country in Africa to report the emergence of this highly pathogenic virus. Here we analyse H5N1 sequences in poultry from two different farms in Lagos state and find that three H5N1 lineages were independently introduced through routes that coincide with the flight paths of migratory birds, although independent trade imports cannot be excluded.

The presence of H5N1 avian influenza virus in Africa was first reported by the United Nations Food and Agriculture Organization in commercial poultry farms in Kaduna state in northern Nigeria on 7 February 2006. The poultry farming industry is second only to oil production in Nigeria and is particularly vulnerable to the introduction of infectious agents because chickens are imported from all over the world without rigorous biosecurity safeguards. This vulnerability is increased because the country has many bird sanctuaries along the two flight paths that link Nigeria with the southern Russian region and Europe, and with western Asia<sup>1</sup>. However, a sero-epidemiological study<sup>2</sup> of 65 flocks in the southwestern states that form the hub of the Nigerian poultry industry showed no evidence of avian influenza in poultry up to the time the study was terminated in May 2004.

We have now recovered virus isolates that tested positive in a generic test<sup>3</sup> for influenza A from 18 cloacal swabs taken from poultry from two farms ('SO' on 24 February 2006 and 'BA' on 8 March 2006) in Lagos state in the far southwest of the country (for map and methods, see supplementary information). Sequencing of these isolates confirmed the presence of H5N1. The viral haemagglutinin cleavage-site sequence PQGERRRKKRG (single-letter amino-acid notation) was identical to that of the highly pathogenic strains already found in Europe, Russia and central Asia. Figure 1 shows that representative sequences from the complete haemagglutinin gene cluster with those of the 2005–06 H5N1



**Figure 1 | H5N1 in Nigerian poultry.** Phylogeny of the viral haemagglutinin gene in the Nigerian lineages and in the most closely related strains. The maximum likelihood method was used, with estimated transition/transversion ratio and a gamma-distribution substitution model and 100 bootstrap values. Bootstrap values of the principal nodes are shown. A/chicken/Hebei/2005 was used as an outgroup to build the tree. Countries having their first outbreak in 2006 are shown in blue; west Asian and Russian strains are in yellow and orange, respectively; southeast Asian strains are in green; and Nigerian strains are in red.

strains from western Asia (Mongolia; Qinghai in China; Crimea in the Ukraine; Kurgan, Astrakhan and Novosibirsk in Russia; Iran; Iraq), Europe (Turkey; Croatia; France; Italy; Germany) and Africa (Egypt).

The percentage variation in nucleotide sequence between the strains from the two Lagos state farms in the southwest and DQ406728, the strain from northern Nigeria, was 0.79%, 0.87% and 1.14%, respectively. The complete haemagglutinin sequences from the BA, SO and northern Nigerian isolates resembled most closely the A/Cygnus olor/Astrakhan/Ast05-2/2005, A/chicken/Egypt/960N3-004/2006 and A/chicken/Kurgan/3/2005 strains, respectively (Fig. 1).

The BA sequences were even more closely related to sequences from swans from Poland and Germany and a buzzard from Denmark (0.3% diversity on the haemagglutinin gene; Ian Brown, personal communication) and, despite the relative proximity of the SO and BA farms (less than 50 kilometres apart), strains from the BA farm were more closely related to the northern Nigerian strain than to the SO farm

strains (see supplementary information). We conclude that the highly pathogenic avian influenza virus strains in northern Nigeria and in Lagos state belong to three distinct genetic lineages and were therefore probably introduced independently. Our sequence data rule out a southeast-Asian origin for these different lineages.

By 20 February 2006, the Nigerian federal government had set up a biosafety zone to contain the virus in the north, whose efficiency was questioned when the virus was detected in Lagos state. Our results indicate, however, that the viruses in the southwest were not introduced from the north, but that instead they all arrived independently, perhaps at inland waters and key bird areas or through unprotected trade. Meanwhile, the H5N1 virus has reached 14 of the 31 federal states of Nigeria and it is difficult to see how the virus might still be contained

without depopulation measures, combined with surveillance and large-scale vaccination.

**M. F. Ducatez\*, C. M. Olinger\*, A. A. Owode†, S. De Landtsheer\*, W. Ammerlaan\*, H. G. M. Niesters‡, A. D. M. E. Osterhaus‡, R. A. M. Fouchier‡, C. P. Muller\***

\*Institute of Immunology, National Public Health Laboratory, 1950 Luxembourg  
e-mail: claude.muller@lns.etat.lu

†Department of Veterinary Medicine, University of Ibadan, Ibadan, Nigeria

‡Department of Virology, Erasmus Medical Center, 3015 GE Rotterdam, The Netherlands

- Olsen, B. et al. *Science* **312**, 384–388 (2006).
- Owode, A. A., Ducatez, M. F. & Muller, C. P. *Avian Dis.* **50**, 222–227 (2006).
- Munster, V. J. et al. *Emerg. Infect. Dis.* **11**, 1545–1551 (2005).

Supplementary information accompanies this communication on Nature's website.

Received 26 May; accepted 13 June 2006.

Competing financial interests: declared none.  
doi:10.1038/442037a

**BRIEF COMMUNICATIONS ARISING** online  
www.nature.com/bca see Nature contents.

## INTELLIGENCE

## Is there a sex difference in IQ scores?

Arising from: S. Blinkhorn *Nature* 438, 31–32 (2005)

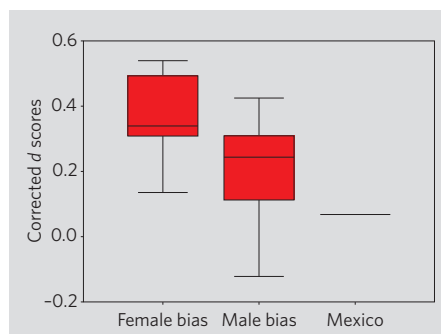
Steve Blinkhorn<sup>1</sup> criticizes our study<sup>2</sup> of samples of university students, in which we found that the average IQ of men is 4.6 points higher than that of women, as measured by the Progressive (or Raven's) Matrices. He maintains that there is a negligible sex difference in adult intelligence. We believe that the principal error of Blinkhorn's criticism is that he does not consider our result in the context of several other studies showing that adult males have an IQ advantage of around 4–6 IQ points<sup>3–7</sup>.

The strongest evidence for our position consists of a meta-analysis<sup>3</sup> derived from 57 general-population samples, many of which are normative or carefully constructed representative samples, with a total of 80,928 participants. This reveals no sex difference in general cognitive ability up to the age of 14 and a significant sex difference at 15, which then increases to its adult value of 5 IQ points in favour of males. There is further evidence for a mean male advantage of 4–6 IQ points in four independent adult samples<sup>5–7</sup> ( $n = 11,896$ ); Jackson and Rushton<sup>8</sup>, in a huge standardization sample ( $n = 102,515$ ) that is much bigger than the Mexico study<sup>9</sup>, also reported a male advantage of 3.6 IQ points among 17-year-olds, which is somewhat greater than our estimates<sup>3,4</sup> for this age group.

At two points in his critique<sup>1</sup>, Blinkhorn addresses the 'file drawer' problem, in which non-significant results fail to get published. Such a bias cannot be operating in this instance because, as Blinkhorn himself notes, almost none of the published literature on the Progressive Matrices has focused on sex differences. Furthermore, the idea that a null sex difference is unpublishable is belied by the huge number of books and articles, including Blinkhorn's, that claim exactly that. Therefore, the fact that 21 out of 22 student samples showed a male advantage suggests the phenomenon is robust.

Blinkhorn criticizes us for not adopting the principle of weighting results by sample size, and for excluding the very large study from Mexico<sup>9</sup>. This misses a central point of meta-analysis. We carried out a number of tests for moderator variables (factors that cause under- or overestimates of the sex difference) and found strong evidence for two: these were the type of test and the tendency of some universities selectively to recruit either brighter men or brighter women. In the presence of strong moderators, many of the studies in the sample provide biased estimates of the sex difference in IQ score.

It is clear from the box plot (Fig. 1) that the Mexico results conform to estimates from the



**Figure 1 | Selection biases in IQ measurements.**

A comparison of Mexico, on the  $d$ -score (male mean minus the female mean, divided by the within-group standard deviation) difference in IQ, with samples showing pro-female and pro-male selection biases among university students. Horizontal bars represent the median; red rectangles, the interquartile range; tails, outliers. Sample sizes: female-biased studies, number of samples  $k = 10$ ,  $n = 6,812$ ; male-biased studies,  $k = 10$ ,  $n = 4,296$ ; Mexico study,  $k = 1$ ,  $n = 9,048$ .

most male-biased samples, which provide substantial underestimates of the sex difference in IQ. Given the strong probability of bias in this sample, to weight it by its sample size (9,048) would risk a serious underestimate of the population sex difference in IQ. For this reason, we followed the advice of a definitive article on meta-analysis<sup>10</sup> and took the median of estimates, including Mexico<sup>9</sup>, which equated to 4.6 IQ points.

Many of Blinkhorn's difficulties stem from his assumption that our focus was on university students. This makes little sense, because the IQ difference in students is dependent on which population is considered, whereas the sex difference in the general population, our actual focus of interest, is highly stable<sup>3</sup>. This also explains our choice of Cohen's  $d$  (Fig. 1),

as this provides a standard metric that partly controls for range restriction. According to accepted meta-analytical procedures<sup>11</sup>, a difference of  $0.31d$  translates into a 4.6-point IQ difference in the general population: this is an underestimate and not the overestimate suggested by Blinkhorn.

Blinkhorn also implies that the Progressive Matrices may be biased against women. This issue has been investigated<sup>12</sup> by using the conservative procedure of eliminating possibly biased items, identified by methods of differential item functioning: that analysis showed negligible evidence of bias.

However, we agree with Blinkhorn on the need for large representative samples, which is why we continue<sup>3,4</sup> to make use of these.

Paul Irwing\*, Richard Lynn†

\*Manchester Business School, University of Manchester, Booth Street West, Manchester M15 6PB, UK  
e-mail: paul.irwing@man.ac.uk

†School of Psychology, University of Ulster, Coleraine, Northern Ireland BT52 1SA, UK

1. Blinkhorn, S. *Nature* **438**, 31–32 (2005).
2. Irwing, P. & Lynn, R. *Br. J. Psychol.* **96**, 505–524 (2005).
3. Lynn, R. & Irwing, P. *Intelligence* **32**, 481–498 (2004).
4. Lynn, R., Allick, J. & Irwing, P. *Intelligence* **32**, 411–424 (2004).
5. Colom, R., Juan-Espinoso, M., Abad, F. J. & Garcia, L. F. *Intelligence* **28**, 57–68 (2000).
6. Colom, R., Garcia, L. F., Juan-Espinoso, M. & Abad, F. J. *Span. J. Psychol.* **5**, 29–35 (2002).
7. Nyborg, H. in *The Scientific Study of General Intelligence* (ed. Nyborg, H.) 187–227 (Elsevier, Amsterdam, 2003).
8. Jackson, D. N. & Rushton, J. P. *Intelligence* (in the press).
9. Backhoff-Escudero, E. *Rev. Mex. Psicol.* **13**, 21–28 (1996).
10. Rosenthal, R. *Psych. Bull.* **118**, 183–192 (1995).
11. Hunter, J. E. & Schmidt, F. L. *Methods of Meta-analysis: Correcting Error and Bias in Research Findings* 2nd edn (Sage, London, 2004).
12. Abad, F. J., Colom, R., Rebollo, I. & Escorial, S. *Pers. Ind. Diff.* **36**, 1459–1470 (2004).

doi:10.1038/nature04966

## INTELLIGENCE

## Blinkhorn replies

Replying to: P Irwing & R. Lynn *Nature* 442, doi:10.1038/nature04966 (2006)

Irwing and Lynn explicitly claim to have identified a sex difference in IQ among university students<sup>1</sup>. They now counter<sup>2</sup> my criticism<sup>3</sup> of the flawed methodology upon which their conclusions are based and claim that an estimate of what this difference may actually be is irrelevant.

Irwing and Lynn's other work on the Progressive Matrices<sup>4</sup> contains no actual statistics,

only estimates of "effect sizes", so it is not susceptible to the ordinary processes of scrutiny. The large study in press<sup>5</sup> is not based on Raven's Matrices but on the American SAT college-entrance examination, has an unexplained excess of female participants, and raises several technical and substantive issues beyond the scope of my original critique.

Their response<sup>2</sup> to my critique, in which they



protest that they are complying with good meta-analytical practice, deflects attention from the inappropriate meta-analytical manoeuvres in which they have become entangled.

For the Standard Progressive Matrices, for instance, they report standard deviations from 3.02 to 9.69, and mean scores from 14.77 to 55.26, with no explanation for these extraordinary variations. They equate each of these widely differing standard deviations in turn to the conventional figure for IQ in the whole population, and then calculate the median of the resulting sex differences, each now in a different, exaggerated and artificial metric. Their meta-analytic paraphernalia are not needed, neither are they helpful, as the original scores themselves provided a suitable metric.

The Mexican study<sup>6</sup> is not biased: rather, Irwing and Lynn's Fig. 1 shows how far removed they have become from the data. Their argument here is circular: the sex difference is vanishingly small compared with their sample of smaller, less representative groups, so therefore there must be a bias.

In fact, the Mexican undergraduates<sup>6</sup> were

tested as part of the admissions process (which adds the crucial element of motivation that tends to eliminate sex differences) at a university that provides for 70% of the demand in its region, covering the whole range of courses. Four-and-a-half times as many people were tested as were tested in the other nine studies using the Standard Progressive Matrices combined.

There were marked IQ differences between gender-stereotyped disciplines (up to about two-thirds of a standard deviation, over four raw-score points), with engineering and physical science disciplines scoring highest, and social sciences and humanities lowest. But the difference between the sexes was only one-tenth as much, and described by the author as tiny ("minima")<sup>6</sup>.

Most large-scale ability testing goes unreported in academic journals. Those who engage with recruitment and selection campaigns rarely have the need, the motivation, or indeed (as in my case) the permission to report all they find. That is why there is a big file-drawer effect. The place to look for group differences is in the archives of institutional

test users, where one can also find the results of efforts to identify and eliminate possible ethnic and sex biases. That, and not a suggestion that Raven's Matrices are biased, was the import of my earlier remarks<sup>3</sup>.

I do not say that Irwing and Lynn's approach to meta-analysis is idiosyncratic, rather that their use of it here is methodologically inappropriate, statistically bizarre and unsuited to the estimation of subpopulation parameters. It has wrongly given rise to claims that IQ tests are biased against women and so should be abandoned.

#### Steve Blinkhorn

Psychometric Research & Development,  
Brewmaster House, The Maltings,  
St Albans AL1 3HT, UK  
e-mail: steve@prd.co.uk

1. Irwing, P. & Lynn, R. *Br. J. Psychol.* **96**, 505–524 (2005).
2. Irwing, P. & Lynn, R. *Nature* **442**, doi:10.1038/nature04967 (2006).
3. Blinkhorn, S. *Nature* **438**, 31–32 (2005).
4. Lynn, R. & Irwing, P. *Intelligence* **32**, 481–498 (2004).
5. Jackson, D. N. & Rushton, J. P. *Intelligence* (in the press).
6. Backhoff-Escudero, E. *Rev. Mex. Psicol.* **13**, 21–28 (1996).

doi:10.1038/nature04967

# Intracellular pattern recognition receptors in the host response

Etienne Meylan<sup>1</sup>, Jürg Tschopp<sup>1</sup> & Michael Karin<sup>2</sup>

**The innate immune system relies on its capacity to rapidly detect invading pathogenic microbes as foreign and eliminate them. Indeed, Toll-like receptors are a class of membrane receptors that sense extracellular microbes and trigger anti-pathogen signalling cascades. Recently, intracellular microbial sensors have also been identified, including NOD-like receptors and the helicase-domain-containing antiviral proteins RIG-I and MDA5. Some of these cytoplasmic molecules sense microbial, as well as non-microbial, danger signals, but the mechanisms of recognition used by these sensors remain poorly understood. Nonetheless, it is apparent that these proteins are likely to have critical roles in health and disease.**

In an environment rich in potentially harmful microbes, survival of the host organism depends on rapidly mounted defence responses. Such protective mechanisms are evolutionarily conserved in all multicellular organisms and are collectively referred to as innate immunity<sup>1</sup>. Unlike adaptive immunity, which is based on millions of lymphoid cell-surface receptors (generated by complex gene rearrangements) that recognize an infinite variety of antigens, the innate immune system is based on a much smaller number of receptors, called pattern recognition receptors (PRRs). For the most part, PRRs recognize conserved molecular patterns that distinguish foreign organisms—viruses, bacteria, fungi and parasites—from cells of their hosts<sup>1</sup>. Such pathogen-associated molecular patterns (PAMPs) include viral nucleic acids, components of bacterial and fungal cell walls, flagellar proteins, and more. However, this detection system is not foolproof and it can also be activated by a variety of normal host proteins and danger signals that are released by dying cells<sup>2</sup>. Inadvertent activation of PRRs by normal host molecules may have an important role in the pathogenesis of various autoimmune and inflammatory diseases<sup>2</sup>.

The first family of PRRs studied in detail was the Toll-like receptor (TLR) family<sup>3</sup>. TLRs are classical transmembrane proteins whose ligand-binding domains—composed of leucine-rich repeats (LRRs)—point towards the extracellular milieu or to the topologically equivalent lumen of membrane-enclosed intracellular compartments<sup>4</sup>. Thus, TLRs recognize either extracellular or membrane-encased foreign organisms. In either case, the TLR signal transduction domains, known as Toll/IL-1 receptor (TIR) domains, are directed towards the cytoplasm and transduce their signals through interaction with cytoplasmic adaptor proteins, which also contain TIR domains<sup>4</sup>. Ligand-induced receptor and adaptor dimerization results in recruitment and activation of additional signalling proteins and eventual triggering of a host-defence response, entailing production of antimicrobial peptides, proinflammatory chemokines, and cytokines, antiviral cytokines, adhesion molecules and enzymes that catalyse production of secondary inflammatory mediators and bactericidal molecules. Recently, however, two other families of PRRs were described: the NLRs (NOD-like receptors) and the RLHs (RIG-like helicases). Unlike TLRs, these families consist of soluble proteins that survey the cytoplasm for signs that broadcast the presence of intracellular invaders. This review will focus on recent progress in

understanding the function and signalling mechanisms used by cytoplasmic innate immune receptors.

## NOD-like receptors, or NLRs

The first intracellular PAMP sensors to be discovered were NOD1 and NOD2 (also known as CARD4 and CARD15, respectively; refs 5, 6). They are members of the emerging NLR family (also known as NOD-LRR, NACHT-LRR and CATERPILLER) implicated in such host defence<sup>7–9</sup>. NLRs contain three distinct domains: an amino-terminal CARD or pyrin effector domain (with the exception of NAIP and possibly NOD5), a nucleotide-binding and oligomerization domain (the so-called ‘NACHT’ domain), and a variable number of carboxy-terminal LRRs<sup>9</sup> (Fig. 1). NLRs comprise two large sub-classes—the NODs (NOD1–5) and 14 members of the NALP clan<sup>9</sup>. With the IPAF, NAIP and CIITA group of proteins, a total number of 22 NLR members have been identified so far<sup>9</sup>. The biological significance of most of the NLRs remains to be determined.

NLR activation is suggested to be similar to the presumed mechanism described for the structurally related plant resistance genes<sup>10</sup> or Apaf-1 (ref. 11). NLRs are synthesized in an autorepressed, inactive form. Their LRRs are thought to fold back onto the remainder of the protein, thereby blocking NACHT-mediated oligomerization. Specific activation of the sensor platform is triggered through direct or indirect binding of ligands to the LRRs, thereby inducing NACHT-dependent oligomerization and formation of higher-order complexes. Rather than repeat several recent reviews<sup>7–9</sup>, the next sections will focus mainly on the function of two NLRs: NOD2 and NALP3 (also called cryopyrin).

## The NOD proteins

NOD1 and NOD2 detect bacterial peptidoglycan, and through a CARD-dependent recruitment of RIP2 (refs 12, 13) drive activation of mitogen-activated protein kinases (MAPKs) and NF- $\kappa$ B (Fig. 2). The minimal structural requirement for peptidoglycan-mediated activation of NOD2 is muramyl dipeptide, which is present in both Gram-positive and Gram-negative bacteria<sup>12,14</sup>. Muramyl dipeptide can be released through the action of bacterial hydrolases during cell-wall biosynthesis and remodelling, or after degradation of ingested bacteria by host lysozyme. Exactly how muramyl dipeptide

<sup>1</sup>Department of Biochemistry, University of Lausanne, Chemin des Boveresses 155, CH-1066 Epalinges, Switzerland. <sup>2</sup>Laboratory of Gene Regulation and Signal Transduction, Department of Pharmacology, School of Medicine, University of California at San Diego, 9500 Gilman Drive, La Jolla, California 92093-0723, USA.

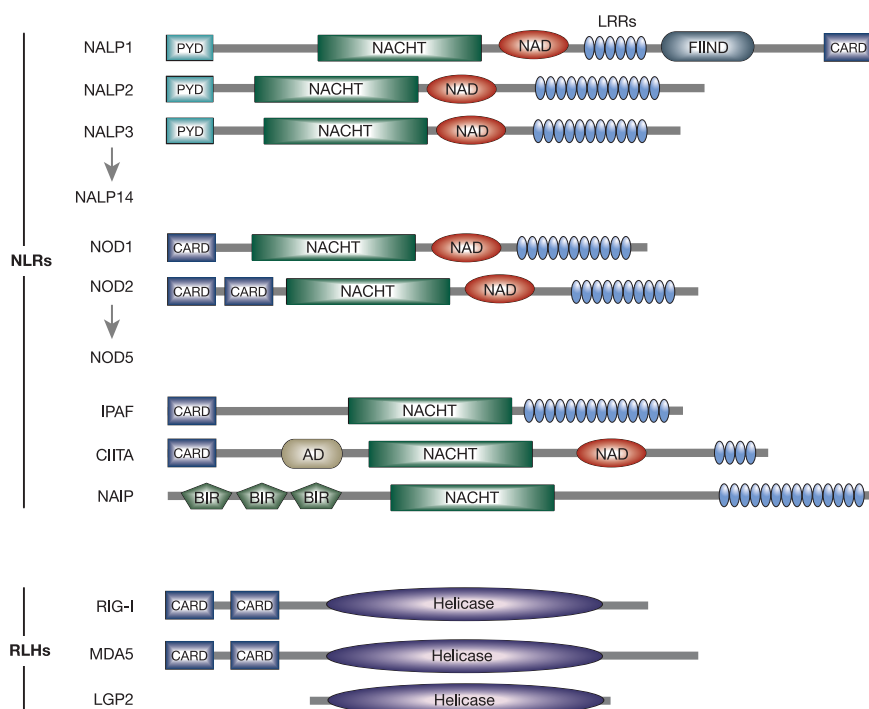


finds its way into the cytoplasm is not clear, but a specific transport system may be involved<sup>15</sup>. It is also unknown whether muramyl dipeptide is recognized directly by NOD2 or by another protein that interacts with NOD2. Although the LRRs of NOD2 are required for muramyl-dipeptide-induced signalling<sup>16</sup>, it should be noted that direct binding of muramyl dipeptide to NOD2 is yet to be demonstrated. Furthermore, cellular muramyl-dipeptide-binding activities differ in their molecular weight from NOD2 (ref. 7). Thus, NOD2 may use its LRR to interact with the actual muramyl dipeptide receptor. There is evidence that NOD2 mediates innate immunity to *Streptococcus pneumoniae*<sup>17</sup>, *Mycobacteria*<sup>18</sup> and *Listeria monocytogenes*<sup>19</sup>. Increased susceptibility to *Listeria* is linked to the impaired expression of a subgroup of intestinal antimicrobial peptides, known as cryptidins, in NOD2-deficient mice<sup>20</sup>. In contrast to NOD2, NOD1 has a more restricted specificity. It detects only meso-diaminopimelic-acid-containing peptidoglycan, which is found in most Gram-negative, but not Gram-positive, bacteria<sup>21</sup>.

Much of the recent interest in NOD2 was generated by the association of *NOD2* mutations with two chronic inflammatory disorders: Crohn's disease and Blau syndrome. Crohn's disease is an inflammatory bowel disease characterized by granulomatous inflammation of the distal ileum, associated with increased production of IL-1 $\beta$  and TNF $\alpha$ , and activation of NF- $\kappa$ B, in lamina propria inflammatory cells<sup>22,23</sup>. Almost 50% of familial Crohn's disease cases in Western populations are linked to three *NOD2* mutant alleles: R702W, G908R and L1007fsinsC (refs 24, 25). Homozygosity for either allele increases the risk of Crohn's disease by up to 40-fold, but many patients who are homozygotes for these mutations remain healthy, and even those who develop the disease are asymptomatic for the first 10–15 years of their life<sup>26</sup>. Thus, it is likely that additional factors, both genetic and environmental, act in conjunction with *NOD2* mutations<sup>22</sup>. Two of the mutant alleles—

R702W and G908R—specify single amino-acid substitutions within the LRR, whereas the third—L1007fsinsC—results in a frame-shift mutation and removes the last 33 amino acids of the NOD2 polypeptide. Transient expression of NOD2 in non-myeloid cells confers the ability to activate a co-transfected NF- $\kappa$ B reporter construct by exogenous muramyl dipeptide, and using this assay it was concluded that all three Crohn's-disease-linked mutations impair muramyl dipeptide recognition<sup>7</sup>. In addition, cultured macrophages or peripheral blood monocytes from Crohn's disease patients that are homozygous for the *NOD2* mutant alleles also show defects in activation of NF- $\kappa$ B and induction of proinflammatory genes in response to muramyl dipeptide<sup>12,27</sup>. However, these results, which suggest that Crohn's-disease-linked *NOD2* mutations are loss-of-function mutations, are not entirely consistent with the common clinical observations of elevated proinflammatory cytokines and presence of activated NF- $\kappa$ B in the lamina propria of Crohn's disease patients<sup>22,23</sup>. Indeed, the analysis of dendritic cells derived from such patients revealed increased expression of certain inflammatory marker genes including IL-1 $\beta$  upon incubation with muramyl dipeptide<sup>78</sup>.

In an attempt to elucidate the functional relationship between *NOD2* mutations and Crohn's disease, two different mouse *Nod2* mutants were generated: one a complete knockout, which no longer expresses the polypeptide<sup>20</sup>, and the other a knockin mutant, *Nod2*<sup>2939ic</sup>, which carries the same mutation as the human L1007fsinsC allele<sup>28</sup>. The latter mouse expresses a normal amount of a truncated NOD2 missing the last 33 amino acids. Both mouse mutants are asymptomatic and do not develop a spontaneous disease. Consistent with the complete absence of NOD2, macrophages from *Nod2*<sup>-/-</sup> mice do not produce proinflammatory cytokines or undergo NF- $\kappa$ B activation after muramyl dipeptide challenge<sup>20</sup>. Conversely, *Nod2*<sup>2939ic</sup> macrophages show elevated



**Figure 1 | Intracellular sensors of bacteria, viruses and danger signals.**

NOD-like receptors (NLRs) are characterized by three distinct domains: the putative ligand-sensing leucine-rich repeats (LRRs); the NACHT domain, which mediates oligomerization; and an effector domain, which can be a pyrin domain (PYD), a CARD (caspase recruitment domain) or a BIR (baculovirus IAP repeat) domain. Most of the NLRs also contain a NACHT-associated domain (NAD). NLRs comprise two large subfamilies:

14 members of the PYD-containing NALP clan and five members of the mostly CARD-containing NODs (NOD5 does not contain a typical CARD domain). The two CARD-containing proteins CIITA and IPAF, and the BIR-containing NAIP protein, constitute the remaining NLR members. RIG-like helicases (RLHs) contain a helicase and two CARD domains, except for LGP2 where the CARDs are absent. Additional abbreviations: FIIND, function to find; AD, activation domain.

responsiveness to muramyl dipeptide, measured by either NF- $\kappa$ B and IKK (I $\kappa$ B kinase) activation or induction of proinflammatory cytokines<sup>28</sup>. However, even in *Nod2*<sup>2939ic</sup> macrophages the ability of muramyl dipeptide to elicit NF- $\kappa$ B activation or induce the expression of genes encoding proinflammatory cytokines is modest at best, relative to the activity of typical TLR ligands such as LPS (lipopolysaccharide; ligand for TLR4) or polyinosinic:polycytidylic acid (poly(I:C); ligand for TLR3). Given that macrophages and dendritic cells, which specialize in the recognition of pathogens, express many distinct TLRs capable of strong NF- $\kappa$ B activation, it is rather unlikely that the primary signalling function of NOD2 (and probably other NOD proteins) is also NF- $\kappa$ B activation. In other words, given the presence of TLRs that recognize many different microbial PAMPs, it is improbable that the complete absence of NOD2 will generate a 'hole' in the NF- $\kappa$ B activation response, one of the principal effector arms of the host defence system. Notably, macrophages from *Nod2*<sup>2939ic</sup> mice also show increased IL-1 $\beta$  secretion on muramyl dipeptide stimulation<sup>28</sup>. Furthermore, dendritic cells from Crohn's disease patients that are homozygous for the same mutation show elevated expression of IL-1 $\beta$  mRNA after stimulation with muramyl dipeptide<sup>78</sup>. The results contrast with another report claiming lack of muramyl dipeptide responsiveness in blood monocytes isolated from Crohn's disease patients bearing similar NOD2 mutations<sup>27</sup>. The basis for this discrepancy is yet to be determined.

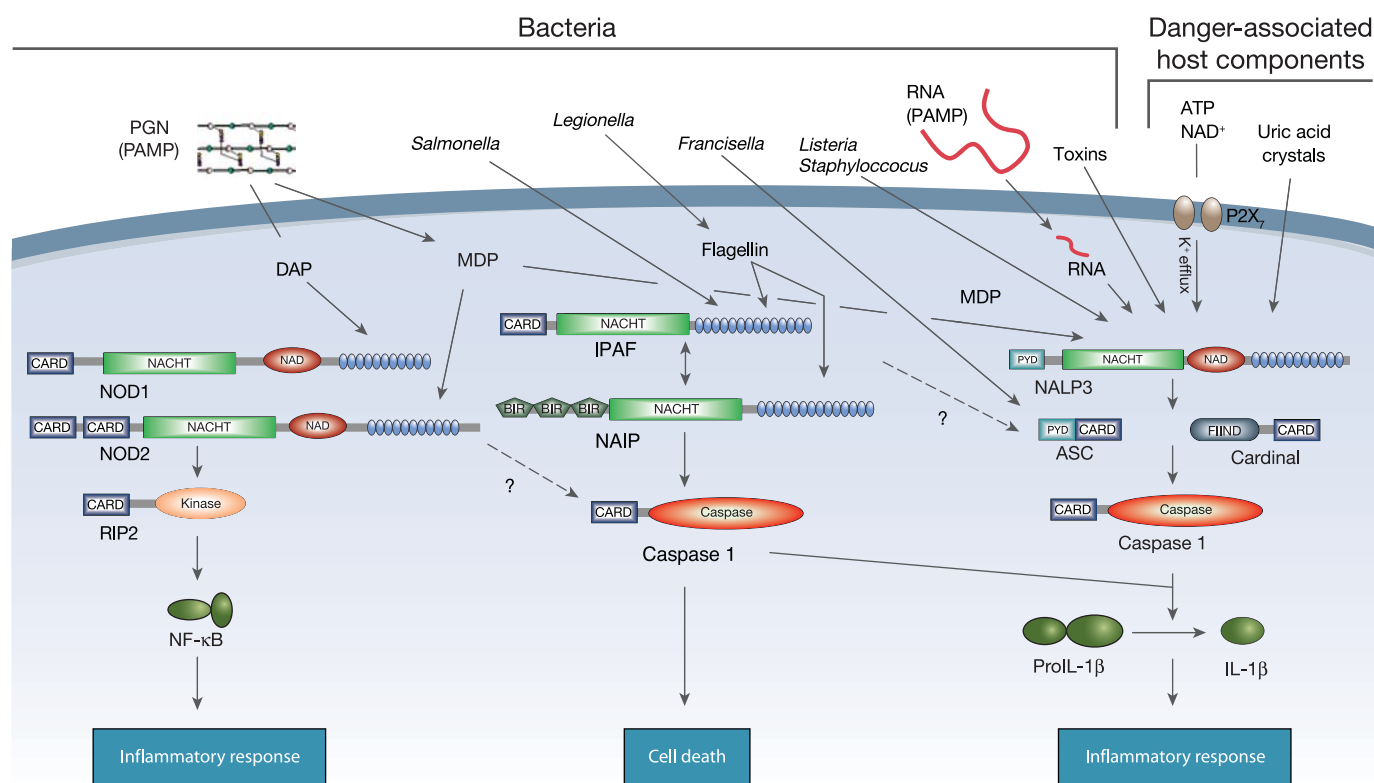
### The NALP3 inflammasome

NALP proteins are characterized by the presence of a N-terminal pyrin effector domain (Fig. 1; for reviews, see refs 7–9). At least some

of them have an important role in activation of proinflammatory caspases through formation of a complex called the 'inflammasome'. Two types of NALP inflammasomes were identified<sup>29</sup>. The NALP1 inflammasome composed of NALP1, the adaptor protein ASC, caspase 1 and caspase 5 (ref. 30), and the NALP2/NALP3 inflammasome that contains—in addition to NALP2 or NALP3—CARDINAL, ASC and caspase 1, but not caspase 5 (ref. 31; Fig. 2). Both inflammasomes control the processing and activation of the proinflammatory cytokines IL-1 $\beta$  and IL-18.

What are the signals activating the NALP3 inflammasome? It seems that not only PAMPs but also the presence of host danger signals (danger-associated molecular patterns, or DAMPs) can be sensed by inflammasomes. Low concentrations of intracellular potassium ( $[K^+] < 70$  mM) seems to be the common denominator of an acute cellular stress triggering inflammasome activation<sup>30</sup>. For example, massive  $K^+$  efflux occurs when high concentrations of ATP (5 mM) activate the purinergic P2X<sub>7</sub> receptor<sup>32</sup>. Potassium-channel-forming bacterial toxins such as *Staphylococcus aureus*  $\alpha$ -toxin or gramicidin from *Bacillus brevis* also activate the inflammasome in a  $K^+$ -dependent manner<sup>32,33</sup>. Thus, the inflammasome does not only sense the presence of bacterial components (see below), but is also endowed with the ability to detect non-microbial danger signals.

In 1994, Matzinger proposed a new concept suggesting that endogenous factors released from tissues undergoing destruction alert the immune system through direct activation of dendritic cells without exposure to foreign substances<sup>34</sup>. Several groups subsequently showed that necrotic cell lysates are a source of endogenous factors that induce dendritic cell activation. One of the components that meets the criteria of an endogenous danger signal is crystalline



**Figure 2 | Activation of NLRs by bacterial and host-derived components.** On recognition of peptidoglycan (PGN)-derived molecules (meso-diaminopimelic acid (DAP) and muramyl dipeptide (MDP), respectively), NOD1 and NOD2 recruit RIP2, which, in turn, activates NF- $\kappa$ B. NOD2 activation can lead to caspase 1 activation through a mechanism that remains to be identified. Muramyl dipeptide (and bacterial RNA) also activates the NALP3 inflammasome, which is formed by NALP3, Cardinal,

ASC and caspase 1, resulting in the processing of proIL-1 $\beta$ . Endogenous danger signals that activate NALP3 include uric acid crystals and the efflux of  $K^+$ , triggered by ATP, NAD<sup>+</sup> or bacterial toxins. Various bacteria trigger activation of distinct NLR pathways. Whereas caspase 1 activation by *Salmonella* and *Legionella* requires IPAF, NAIP (and ASC), but not NALP3, *Listeria* and *Staphylococcus* triggers caspase 1 activation via the NALP3 inflammasome.



monosodium urate<sup>35</sup>. Uric acid is normally present in tissues as a degradation product of purines. Under stress conditions, monosodium urate may be released to eventually form crystals that alert dendritic cells and other immune cells such as macrophages. Although the adjuvant effect of monosodium urate can be advantageous by triggering a potent innate immune response against pathogens, it can also trigger deleterious autoinflammatory responses. Indeed, monosodium urate is recognized as the aetiological agent of gout, a condition in which elevated uric acid is associated with the onset of arthritis. Notably, not only monosodium urate but also calcium pyrophosphate dihydrate crystals—involved in the arthropathy pseudogout—were recently shown to activate the NALP3 inflammasome<sup>36</sup>. Despite the existence of 14 NALPs, all host danger signals tested so far, including those that lower intracellular  $K^+$  concentration, seem to activate caspase 1 through the NALP3 inflammasome<sup>36–38</sup>. This suggests that NALP3 may represent one of the long-sought host danger receptors.

In addition to DAMPs, NALP3 also senses bacterial PAMPs. Interestingly, NALP3 was proposed to detect the same PAMPs—namely peptidoglycan and its degradation product muramyl dipeptide<sup>39</sup>—that activate NOD2, suggesting that these bacterial cell wall components are predominantly sensed by the cytoplasmic NLRs and not extracellular TLRs (the reported peptidoglycan–TLR2 interaction was recently questioned<sup>40</sup>). Moreover, bacterial RNA and small antiviral compounds were also proposed to activate the NALP3 inflammasome independently of TLRs<sup>41</sup>, suggesting that multiple PAMPs may activate the NALP3 inflammasome. However, the physiological relevance of muramyl dipeptide and bacterial RNA detection by NALP3 remains to be addressed *in vivo*. It is known that not all bacteria are sensed by the NALP3 inflammasome despite the presence of peptidoglycan in their cell walls. Whereas *L. monocytogenes* and *S. aureus* do not elicit caspase 1 activation in *Nalp3*<sup>−/−</sup> and *Asc*<sup>−/−</sup> mice, caspase 1 activation mediated by *Francisella tularensis* requires ASC but not NALP3 (ref. 37), suggesting that NALP3-independent

sensing occurs. Similarly, *Salmonella typhimurium*-induced inflammatory response and caspase-1-dependent cell death are ASC-dependent, but not NALP3-dependent<sup>37,38</sup>. Intriguingly, in addition to ASC, *Salmonella*-induced IL-1 $\beta$  processing and cell death also require IPAF<sup>42</sup>—a member of the NLR family that was shown to be activated by intracellular bacterial flagellin<sup>43,44</sup> and that can directly activate caspase 1 without the need of the ASC adaptor. This suggests that *Salmonella* sensing requires a special conduit, provided by IPAF, in order to activate a putative NALPX (ASC-dependent) inflammasome.

Owing to its potentially dangerous proinflammatory action, the synthesis, processing and release of IL-1 $\beta$  by macrophages are tightly controlled and require at least two distinct stimuli<sup>45</sup>. An inflammatory stimulus, given by TLR ligands such as LPS, induces accumulation of large intracellular stores of the 31-kDa proIL-1 $\beta$  (step 1), while a second stimulus activates the inflammasome and caspase 1 (step 2) followed by release of the active mature 17-kDa IL-1 $\beta$  (step 3). The second and third stimuli can be provided by DAMPs such as ATP, NAD<sup>+</sup>, nigericin, bacterial toxins ( $K^+$  efflux) or uric acid crystals<sup>36,41,46,47</sup>. Although TLR activation was proposed to support all three steps<sup>41</sup>, this seems to be quite inefficient as it was not observed by others<sup>32,33,36,37,42</sup>. This 'three-step' model may guarantee that initiation of innate immune defence requires not only a signal from the extracellular milieu via TLRs, but, in addition, a host danger signal, thereby ensuring that innate immunity is not accidentally triggered by, for example, commensal bacteria.

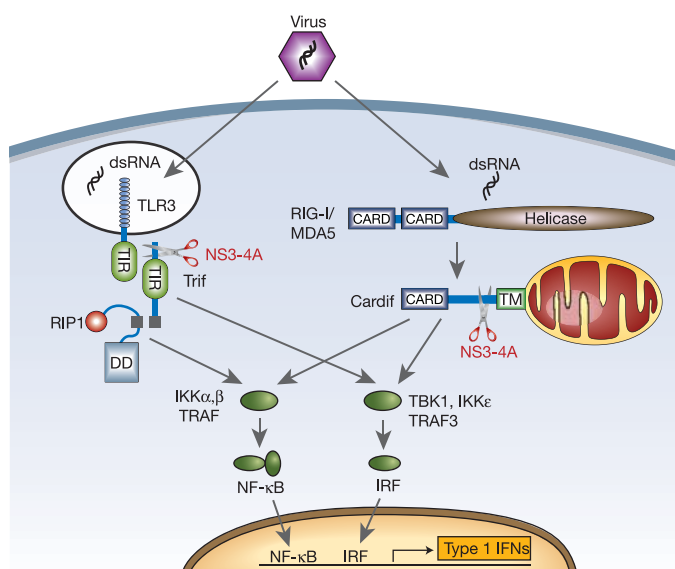
The crucial role of the NALP3 inflammasome has not only been highlighted by analysis of inflammasome-deficient mice<sup>36–38,41,42,48</sup>, but also by the identification of NALP3 mutations associated with three autoinflammatory disorders: Muckle-Wells syndrome, familial cold autoinflammatory syndrome, and neonatal-onset multisystemic inflammatory disease<sup>49,50</sup>. These disorders are associated with constitutive release of IL-1 $\beta$  from monocytes of patients, suggesting that the mutations trigger spontaneous NALP3 oligomerization without the need of a ligand<sup>51</sup>. Importantly, patients with NALP3 mutations are now being treated with the IL-1R antagonist IL1ra (anakinra)<sup>51</sup>. The treatment leads to an almost immediate relief from all symptoms including rashes, periodic fever and arthritis<sup>52–54</sup>.

Little is known about the role of most of the remaining NLRs in innate immunity, as the corresponding knockout mice are not yet available. One exception is NAIP, where genetic studies in mice revealed that NAIP5 is involved in caspase-1-dependent susceptibility of macrophages to *Legionella pneumophila*<sup>55</sup> through sensing of flagellin<sup>56–58</sup>. IPAF is also required for *L. pneumophila* growth restriction<sup>57</sup>, which is explainable by the observation that NAIP5 and IPAF physically interact<sup>57</sup>. *Legionella*-induced IL-1 $\beta$  release, but not bacterial growth, is restricted in ASC-deficient macrophages<sup>42,57</sup>, suggesting that NAIP has an antibacterial role that is independent of the NALP inflammasome.

### RIG-like helicases, or RLHs

Innate immune responses against invading viruses also rely on detection of viral PAMPs and subsequent production of antiviral cytokines such as type-I interferons (IFNs). One prototypical viral PAMP is double-stranded (ds)RNA, which can be detected by TLR3 (ref. 59). Although TLR3 was the first reported dsRNA receptor capable of signalling to IRF and NF- $\kappa$ B—two essential transcription factors that regulate type-I IFN production (Fig. 3)—its role as the primary antiviral receptor was recently challenged. *In vivo* antiviral responses to several viruses, including vesicular stomatitis virus (VSV), lymphocytic choriomeningitis virus (LCMV), reovirus, and murine cytomegalovirus (MCMV), were found to be similar between TLR3-deficient and wild-type mice<sup>60</sup>. Even *in vitro* responses to synthetic dsRNA are rendered TLR3-independent once cells are incubated with sufficient levels of synthetic dsRNA<sup>59</sup>.

Hence, other crucial dsRNA sensors may exist. Recently, three homologous DExD/H box RNA helicases were identified as cyto-



**Figure 3 | Sensing of viral dsRNA by TLR3 and RLH.** Left: On dsRNA binding within endosomes, TLR3 recruits the adaptor Trif through a TIR–TIR interaction. Trif, in turn, recruits RIP1 to activate NF- $\kappa$ B, via TRAFs and the IKK (I $\kappa$ B kinase) complex. Trif also recruits TBK1/IKK $\epsilon$  and TRAF3 to activate IRF3/7. Right: On dsRNA binding, RIG-I (and also MDA5) recruits mitochondrially anchored Cardif, which, in turn, recruits TBK1/IKK $\epsilon$  and TRAF3 to activate NF- $\kappa$ B and IRF, resulting in the induction of type I IFN. During an HCV infection, both Trif and Cardif are cleaved and inactivated by the HCV NS3-4A protease. Additional abbreviations: DD, death domain; TIR, Toll/IL-1 receptor; TM, transmembrane region.

plasmic sensors of virally derived dsRNA (members of the RLHs). Two family members, retinoic-acid-inducible gene I (RIG-I; also called DDX58) and melanoma-differentiation-associated gene 5 (MDA5; also called Helicard), share two N-terminal CARDs followed by an RNA helicase domain (Fig. 1). When activated during a viral infection, dsRNA stimulation or upon overexpression in cells, RIG-I and MDA5 trigger activation of NF- $\kappa$ B and IRF3/7, which cooperate in induction of antiviral type I IFN<sup>61,62</sup> (Fig. 3). The third helicase, LGP2, is devoid of CARDs and prevents activation of antiviral responses, most likely through sequestration of dsRNA from RIG-I or MDA5 (refs 63, 64). Of note, other cytoplasmic dsRNA-binding proteins (such as PKR and 2',5'-oligoadenylate synthetase) also constitute important antiviral molecules, however they act in a second, IFN-dependent phase<sup>65</sup>.

The *in vivo* relevance of RIG-I and MDA5 was addressed through generation of knockout mice, which show impaired antiviral responses against different viruses. RIG-I, but not MDA5, mounts antiviral responses against the positive-strand single-stranded (ss)RNA virus Japanese encephalitis virus, and a set of negative-strand ssRNA viruses such as Newcastle disease virus, VSV, Sendai virus and influenza virus. In contrast, MDA5 senses the presence of the positive-strand ssRNA picornavirus encephalomyocarditis virus<sup>66,67</sup>. Interestingly, although antiviral responses are strongly affected in conventional dendritic cells and mouse embryonic fibroblasts derived from RIG-I-deficient animals, no difference compared to wild-type counterparts is observed in plasmacytoid dendritic cells, which instead require the TIR-containing TLR adaptor MyD88 and probably TLR7 and/or TLR9, which also recognize viral nucleic acids, to mount antiviral responses<sup>66</sup>.

Mechanistically, RIG-I and MDA5 use their CARDs to signal downstream events, which suggest the existence of (a) putative CARD adaptor molecule(s). Such an adaptor was recently identified, and named CARD adaptor inducing IFN- $\beta$  (Cardif; also called MAVS, IPS-1 or VISA)<sup>68–71</sup>. Cardif contains a N-terminal CARD that interacts with both RIG-I and MDA5, whereas its C terminus harbours a transmembrane region that targets Cardif to the outer mitochondrial membrane<sup>69</sup> (Fig. 3). Hence, both RIG-I and MDA5 signalling converge at Cardif, a hypothesis that was confirmed with the recent generation of Cardif-deficient mice<sup>72</sup>. Notably, when Cardif is released from mitochondria to the cytoplasm, or is targeted to another organelle such as the endoplasmic reticulum, it no longer mediates downstream IRF and NF- $\kappa$ B activation<sup>69</sup>. The relevance of Cardif as a crucial antiviral molecule is also inferred from experiments that showed Cardif cleavage and inactivation during hepatitis C virus (HCV) infection<sup>68,73</sup>. Further, the TLR3 adaptor Trif is targeted by the same HCV viral protease (NS3-4A), which also results in its inactivation<sup>74</sup>. Hence, the development of small antiviral compounds targeting NS3-4A, which is also indispensable for HCV maturation, is an attractive strategy for treatment of HCV infections<sup>75</sup>. Once activated by dsRNA, the Cardif platform most probably recruits appropriate signalling intermediates, such as IKKs (namely IKK $\alpha$ ,  $\beta$ ,  $\epsilon$  and TBK1) to activate NF- $\kappa$ B and IRF transcription factors<sup>68–71</sup>. Furthermore, TRAF3, a unique antiviral TRAF molecule implicated in type I IFN production, may also participate in these intracellular events<sup>76,77</sup>. Undoubtedly, future research will aim at understanding the molecular mechanisms of antiviral pathways that emanate from the recently described RLH family, and how viruses counteract them.

### Concluding remarks

A growing number of receptors/sensors for intracellular pathogens and endogenous danger signals are currently being discovered. During microbial infection, these receptors are activated by PAMPs and alert the immune system to take the necessary countermeasures. If infection results in cell damage, endogenous activators are released (such as crystalline uric acid), setting the stage for an optimal immune response. Many questions concerning how these receptors

distinguish between pathogenic and commensal bacteria, and between danger signals and normal cellular constituents, remain unanswered. For example, how does RIG-I undergo activation only when RNA originates from viruses and not from cellular RNA. It seems that the proper transient activation of NLR responses is difficult to sustain and is easily disrupted, leading to various inflammatory diseases. Identification of these novel sensors and their signalling mechanisms may therefore provide us with new targets for the development of drugs for treating such poorly understood diseases. The therapeutic potential that this new insight provides is most dramatically exemplified by the success of IL-1ra in treating hereditary inflammatory disorders.

- Janeway, C. A. Jr & Medzhitov, R. Innate immune recognition. *Annu. Rev. Immunol.* **20**, 197–216 (2002).
- Karin, M., Lawrence, T. & Nizet, V. Innate immunity gone awry: linking microbial infections to chronic inflammation and cancer. *Cell* **124**, 823–835 (2006).
- Medzhitov, R. Toll-like receptors and innate immunity. *Nature Rev. Immunol.* **1**, 135–145 (2001).
- Akira, S. & Takeda, K. Toll-like receptor signalling. *Nature Rev. Immunol.* **4**, 499–511 (2004).
- Inohara, N. *et al.* Nod1, an Apaf-1-like activator of caspase-9 and nuclear factor- $\kappa$ B. *J. Biol. Chem.* **274**, 14560–14567 (1999).
- Ogura, Y. *et al.* Nod2, a Nod1/Apaf-1 family member that is restricted to monocytes and activates NF- $\kappa$ B. *J. Biol. Chem.* **276**, 4812–4818 (2001).
- Inohara, N., Chamailard, M., McDonald, C. & Nunez, G. NOD-LRR proteins: role in host–microbial interactions and inflammatory disease. *Annu. Rev. Biochem.* **74**, 355–383 (2005).
- Ting, J. & Davis, B. CATERPILLER: a novel gene family important in immunity, cell death, and diseases. *Annu. Rev. Immunol.* **23**, 387–414 (2005).
- Martinon, F. & Tschopp, J. NLRs join TLRs as innate sensors of pathogens. *Trends Immunol.* **26**, 447–454 (2005).
- Jones, D. A. & Takemoto, D. Plant innate immunity—direct and indirect recognition of general and specific pathogen-associated molecules. *Curr. Opin. Immunol.* **16**, 48–62 (2004).
- Riedl, S. J., Li, W., Chao, Y., Schwarzenbacher, R. & Shi, Y. Structure of the apoptotic protease-activating factor 1 bound to ADP. *Nature* **434**, 926–933 (2005).
- Inohara, N. *et al.* Host recognition of bacterial muramyl dipeptide mediated through NOD2. Implications for Crohn's disease. *J. Biol. Chem.* **278**, 5509–5512 (2003).
- Viala, J. *et al.* Nod1 responds to peptidoglycan delivered by the *Helicobacter pylori* cag pathogenicity island. *Nature Immunol.* **5**, 1166–1174 (2004).
- Girardin, S. E. *et al.* Nod2 is a general sensor of peptidoglycan through muramyl dipeptide (MDP) detection. *J. Biol. Chem.* **278**, 8869–8872 (2003).
- Vavricka, S. R. *et al.* hPepT1 transports muramyl dipeptide, activating NF- $\kappa$ B and stimulating IL-8 secretion in human colonic Caco2/bbe cells. *Gastroenterology* **127**, 1401–1409 (2004).
- Tanabe, T. *et al.* Regulatory regions and critical residues of NOD2 involved in muramyl dipeptide recognition. *EMBO J.* **23**, 1587–1597 (2004).
- Opitz, B. *et al.* Nucleotide-binding oligomerization domain proteins are innate immune receptors for internalized *Streptococcus pneumoniae*. *J. Biol. Chem.* **279**, 36426–36432 (2004).
- Ferwerda, G. *et al.* NOD2 and toll-like receptors are nonredundant recognition systems of *Mycobacterium tuberculosis*. *PLoS Pathog.* **1**, 279–285 (2005).
- Kanazawa, N. *et al.* Early-onset sarcoidosis and CARD15 mutations with constitutive nuclear factor- $\kappa$ B activation: common genetic etiology with Blau syndrome. *Blood* **105**, 1195–1197 (2005).
- Kobayashi, K. S. *et al.* Nod2-dependent regulation of innate and adaptive immunity in the intestinal tract. *Science* **307**, 731–734 (2005).
- Chamaillard, M. *et al.* An essential role for NOD1 in host recognition of bacterial peptidoglycan containing diaminopimelic acid. *Nature Immunol.* **4**, 702–707 (2003).
- Podolsky, D. K. Inflammatory bowel disease. *N. Engl. J. Med.* **347**, 417–429 (2002).
- McAlindon, M. E., Hawkey, C. J. & Mahida, Y. R. Expression of interleukin 1 $\beta$  and interleukin 1 $\beta$  converting enzyme by intestinal macrophages in health and inflammatory bowel disease. *Gut* **42**, 214–219 (1998).
- Hugot, J. P. *et al.* Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* **411**, 599–603 (2001).
- Ogura, Y. *et al.* A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* **411**, 603–606 (2001).
- Eckmann, L. & Karin, M. NOD2 and Crohn's disease: loss or gain of function? *Immunity* **22**, 661–667 (2005).
- Li, J. *et al.* Regulation of IL-8 and IL-1 $\beta$  expression in Crohn's disease associated NOD2/CARD15 mutations. *Hum. Mol. Genet.* **13**, 1715–1725 (2004).
- Maeda, S. *et al.* Nod2 mutation in Crohn's disease potentiates NF- $\kappa$ B activity and IL-1 $\beta$  processing. *Science* **307**, 734–738 (2005).
- Martinon, F. & Tschopp, J. Inflammatory caspases: linking an intracellular innate immune system to autoinflammatory diseases. *Cell* **117**, 561–574 (2004).

30. Martinon, F., Burns, K. & Tschopp, J. The inflammasome. A molecular platform triggering activation of inflammatory caspases and processing of proIL- $\beta$ . *Mol. Cell* **10**, 417–426 (2002).
31. Agostini, L. *et al.* NALP3 forms an IL-1 $\beta$  processing inflammasome with increased activity in Muckle–Wells auto-inflammatory disorder. *Immunity* **20**, 319–325 (2004).
32. Perregaux, D. & Gabel, C. A. Interleukin-1 $\beta$  maturation and release in response to ATP and nigericin. Evidence that potassium depletion mediated by these agents is a necessary and common feature of their activity. *J. Biol. Chem.* **269**, 15195–15203 (1994).
33. Walev, I., Reske, K., Palmer, M., Valeva, A. & Bhakdi, S. Potassium-inhibited processing of IL-1 $\beta$  in human monocytes. *EMBO J.* **14**, 1607–1614 (1995).
34. Matzinger, P. Tolerance, danger, and the extended family. *Annu. Rev. Immunol.* **12**, 991–1045 (1994).
35. Shi, Y., Evans, J. E. & Rock, K. L. Molecular identification of a danger signal that alerts the immune system to dying cells. *Nature* **425**, 516–521 (2003).
36. Martinon, F., Petrilli, V., Mayor, A., Tardivel, A. & Tschopp, J. Gout-associated uric acid crystals activate the NALP3 inflammasome. *Nature* **440**, 237–241 (2006).
37. Mariathasan, S. *et al.* Cryopyrin activates the inflammasome in response to toxins and ATP. *Nature* **440**, 228–232 (2006).
38. Sutterwala, F. S. *et al.* Critical role for NALP3/CIAS1/cryopyrin in innate and adaptive immunity through its regulation of caspase-1. *Immunity* **24**, 317–327 (2006).
39. Martinon, F., Agostini, L., Meylan, E. & Tschopp, J. Identification of bacterial muramyl dipeptide as activator of the NALP3/cryopyrin inflammasome. *Curr. Biol.* **14**, 1929–1934 (2004).
40. Travassos, L. H. *et al.* Toll-like receptor 2-dependent bacterial sensing does not occur via peptidoglycan recognition. *EMBO Rep.* **5**, 1–7 (2004).
41. Kanneganti, T. D. *et al.* Bacterial RNA and small antiviral compounds activate caspase-1 through cryopyrin/Nalp3. *Nature* **440**, 233–236 (2006).
42. Mariathasan, S. *et al.* Differential activation of the inflammasome by caspase-1 adaptors ASC and Ipaf. *Nature* **430**, 213–218 (2004).
43. Franchi, L. *et al.* Cytosolic flagellin requires Ipaf for activation of caspase-1 and interleukin 1 $\beta$  in *Salmonella*-infected macrophages. *Nature Immunol.* **7**, 576–582 (2006).
44. Miao, E. A. *et al.* Cytoplasmic flagellin activates caspase-1 and secretion of interleukin 1 $\beta$  via Ipaf. *Nature Immunol.* **7**, 569–575 (2006).
45. Dinarello, C. A. Biologic basis for interleukin-1 in disease. *Blood* **87**, 2095–2147 (1996).
46. Cheneval, D. *et al.* Increased mature interleukin-1 $\beta$  (IL-1 $\beta$ ) secretion from THP-1 cells induced by nigericin is a result of activation of p45 IL-1 $\beta$ -converting enzyme processing. *J. Biol. Chem.* **273**, 17846–17851 (1998).
47. Kawamura, H., Aswad, F., Minagawa, M., Govindarajan, S. & Dennert, G. P2X<sub>7</sub> receptors regulate NKT cells in autoimmune hepatitis. *J. Immunol.* **176**, 2152–2160 (2006).
48. Yamamoto, M. *et al.* ASC is essential for LPS-induced activation of procaspase-1 independently of TLR-associated signal adaptor molecules. *Genes Cells* **9**, 1055–1067 (2004).
49. Hoffman, H. M., Mueller, J. L., Broide, D. H., Wanderer, A. A. & Kolodner, R. D. Mutation of a new gene encoding a putative pyrin-like protein causes familial cold autoinflammatory syndrome and Muckle–Wells syndrome. *Nature Genet.* **29**, 301–305 (2001).
50. McDermott, M. F. Genetic clues to understanding periodic fevers, and possible therapies. *Trends Mol. Med.* **8**, 550–554 (2002).
51. Dinarello, C. A. Blocking IL-1 in systemic inflammation. *J. Exp. Med.* **201**, 1355–1357 (2005).
52. Hawkins, P. N., Lachmann, H. J., Aganna, E. & McDermott, M. F. Spectrum of clinical features in Muckle–Wells syndrome and response to anakinra. *Arthritis Rheum.* **50**, 607–612 (2004).
53. Hawkins, P. N., Lachmann, H. J. & McDermott, M. F. Interleukin-1-receptor antagonist in the Muckle–Wells syndrome. *N. Engl. J. Med.* **348**, 2583–2584 (2003).
54. Hoffman, H. M. *et al.* Prevention of cold-associated acute inflammation in familial cold autoinflammatory syndrome by interleukin-1 receptor antagonist. *Lancet* **364**, 1779–1785 (2004).
55. Fortier, A., Diez, E. & Gros, P. Naip5/Birc1e and susceptibility to *Legionella pneumophila*. *Trends Microbiol.* **13**, 328–335 (2005).
56. Molofsky, A. B. *et al.* Cytosolic recognition of flagellin by mouse macrophages restricts *Legionella pneumophila* infection. *J. Exp. Med.* **203**, 1093–1104 (2006).
57. Zamboni, D. S. *et al.* The Birc1e cytosolic pattern-recognition receptor contributes to the detection and control of *Legionella pneumophila* infection. *Nature Immunol.* **7**, 318–325 (2006).
58. Ren, T., Zamboni, D. S., Roy, C. R., Dietrich, W. F. & Vance, R. E. Flagellin-deficient *Legionella* mutants evade caspase-1-and NAIP5-mediated macrophage immunity. *PLoS Pathog.* **2**, 175–183 (2006).
59. Alexopoulou, L., Holt, A. C., Medzhitov, R. & Flavell, R. A. Recognition of double-stranded RNA and activation of NF- $\kappa$ B by Toll-like receptor 3. *Nature* **413**, 732–738 (2001).
60. Edelmann, K. H. *et al.* Does Toll-like receptor 3 play a biological role in virus infections? *Virology* **322**, 231–238 (2004).
61. Yoneyama, M. *et al.* The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. *Nature Immunol.* **5**, 730–737 (2004).
62. Andrejeva, J. *et al.* The V proteins of paramyxoviruses bind the IFN-inducible RNA helicase, mda-5, and inhibit its activation of the IFN- $\beta$  promoter. *Proc. Natl Acad. Sci. USA* **101**, 17264–17269 (2004).
63. Yoneyama, M. *et al.* Shared and unique functions of the DExD/H-box helicases RIG-I, MDA5, and LGP2 in antiviral innate immunity. *J. Immunol.* **175**, 2851–2858 (2005).
64. Rothenfusser, S. *et al.* The RNA helicase Lgp2 inhibits TLR-independent sensing of viral replication by retinoic acid-inducible gene-I. *J. Immunol.* **175**, 5260–5268 (2005).
65. Samuel, C. E. Antiviral actions of interferons. *Clin. Microbiol. Rev.* **14**, 778–809 (2001).
66. Kato, H. *et al.* Cell type-specific involvement of RIG-I in antiviral response. *Immunity* **23**, 19–28 (2005).
67. Kato, H. *et al.* Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses. *Nature* **441**, 101–105 (2006).
68. Meylan, E. *et al.* Cardif is an adaptor protein in the RIG-I antiviral pathway and is targeted by hepatitis C virus. *Nature* **437**, 1167–1172 (2005).
69. Seth, R. B., Sun, L., Ea, C. K. & Chen, Z. J. Identification and characterization of MAVS, a mitochondrial antiviral signaling protein that activates NF- $\kappa$ B and IRF 3. *Cell* **122**, 669–682 (2005).
70. Kawai, T. *et al.* IPS-1, an adaptor triggering RIG-I- and Mda5-mediated type I interferon induction. *Nature Immunol.* **6**, 981–988 (2005).
71. Xu, L. G. *et al.* VISA is an adapter protein required for virus-triggered IFN- $\beta$  signaling. *Mol. Cell* **19**, 727–740 (2005).
72. Sun, Q. *et al.* The specific and essential role of MAVS in antiviral innate immune responses. *Immunity* **24**, 633–642 (2006).
73. Li, X. D., Sun, L., Seth, R. B., Pineda, G. & Chen, Z. J. Hepatitis C virus protease NS3/4A cleaves mitochondrial antiviral signaling protein off the mitochondria to evade innate immunity. *Proc. Natl Acad. Sci. USA* **102**, 17717–17722 (2005).
74. Li, K. *et al.* Immune evasion by hepatitis C virus NS3/4A protease-mediated cleavage of the Toll-like receptor 3 adaptor protein TRIF. *Proc. Natl Acad. Sci. USA* **102**, 2992–2997 (2005).
75. Lamarre, D. *et al.* An NS3 protease inhibitor with antiviral effects in humans infected with hepatitis C virus. *Nature* **426**, 186–189 (2003).
76. Hacker, H. *et al.* Specificity in Toll-like receptor signalling through distinct effector functions of TRAF3 and TRAF6. *Nature* **439**, 204–207 (2006); published online 23 November 2005.
77. Oganessian, G. *et al.* Critical role of TRAF3 in the Toll-like receptor-dependent and -independent antiviral response. *Nature* **439**, 208–211 (2006); published online 23 November 2005.
78. Braat, H. *et al.* Consequence of functional *Nod2* and *Tlr4* mutations on gene transcription in Crohn's disease patients. *J. Mol. Med.* **83**, 601–609 (2005).

**Acknowledgements** We thank H. Everett, D. Muruve and F. Martinon for critical reading of the manuscript. This work was supported by grants from the Swiss National Science Foundation (J.T.), the Commission of Technology and Innovation (J.T.), and the National Institutes of Health (M.K.). M.K. is an American Cancer Society Research Professor.

**Author Information** Reprints and permissions information is available at [npg.nature.com/reprintsandpermissions](http://npg.nature.com/reprintsandpermissions). The authors declare no competing financial interests. Correspondence should be addressed to J.T. ([jurg.tschopp@unil.ch](mailto:jurg.tschopp@unil.ch)) or M.K. ([Karinoffice@ucsd.edu](mailto:Karinoffice@ucsd.edu)).



# Selective elimination of messenger RNA prevents an incidence of untimely meiosis

Yuriko Harigaya<sup>1</sup>, Hirotsugu Tanaka<sup>1†</sup>, Soichiro Yamanaka<sup>1</sup>, Kayoko Tanaka<sup>1</sup>, Yoshinori Watanabe<sup>1†</sup>, Chihiro Tsutsumi<sup>3</sup>, Yuji Chikashige<sup>3</sup>, Yasushi Hiraoka<sup>3</sup>, Akira Yamashita<sup>2</sup> & Masayuki Yamamoto<sup>1,2</sup>

**Much remains unknown about the molecular regulation of meiosis. Here we show that meiosis-specific transcripts are selectively removed if expressed during vegetative growth in fission yeast. These messenger RNAs contain a cis-acting region—which we call the DSR—that confers this removal via binding to a YTH-family protein Mmi1. Loss of Mmi1 function severely impairs cell growth owing to the untimely expression of meiotic transcripts. Microarray analysis reveals that at least a dozen such meiosis-specific transcripts are eliminated by the DSR-Mmi1 system. Mmi1 remains in the form of multiple nuclear foci during vegetative growth. At meiotic prophase these foci precipitate to a single focus, which coincides with the dot formed by the master meiosis-regulator Mei2. A meiotic arrest due to the loss of the Mei2 dot is released by a reduction in Mmi1 activity. We propose that Mei2 turns off the DSR-Mmi1 system by sequestering Mmi1 to the dot and thereby secures stable expression of meiosis-specific transcripts.**

The cell-cycle switch from mitosis to meiosis is accompanied by a striking change in gene expression profiles, as illustrated previously by genome-wide analyses in both the budding yeast *Saccharomyces cerevisiae*<sup>1,2</sup> and the fission yeast *Schizosaccharomyces pombe*<sup>3</sup>. These microbes initiate meiosis in response to nutrient starvation<sup>4,5</sup>. In fission yeast, nutrient starvation stimulates the transcription of the *ste11* gene via the cAMP-dependent protein kinase<sup>6,7</sup>, which encodes a HMG-family transcription factor, Ste11, that is crucial for sexual development<sup>8</sup>. Ste11 activates a large number of genes including *mei2*, which encodes the master regulator of meiosis in fission yeast<sup>9,10</sup>. The gene product of *mei2* is Mei2, an RNA-binding protein that forms a dot structure in the nucleus and this co-localizes with the polyadenylated RNA molecule, meiRNA, that is encoded by the *sme2* gene<sup>9–12</sup>. The Mei2 dot structure has been characterized as a key factor in the promotion of meiosis<sup>11,12</sup>.

The *ste11* mutant of fission yeast is completely defective in meiosis<sup>8</sup>, indicating that transcriptional regulation is essential for the initiation and progression of the meiotic pathway. However, our analysis has now revealed that the level of certain meiosis-specific transcripts is governed also by a cryptic but crucial regulatory mechanism. Here we demonstrate that the transcripts of a number of meiosis-specific genes contain a region that directs their elimination in mitotic cells. A novel RNA-binding protein, Mmi1, is necessary for this elimination. Furthermore, the Mei2 nuclear dot sequesters Mmi1 and thereby allows stable expression of meiosis-specific mRNAs. These unveil a previously uncharacterized mode of regulation during meiosis. A schematic of our main finding is depicted in Supplementary Fig. S1.

## Elimination of meiotic mRNAs in mitosis

The *mei4* transcript, which encodes a meiosis-specific transcription factor of the forkhead family, is abundant during meiosis but undetectable in growing cells<sup>3,13</sup> (Fig. 1a). We measured the *mei4* promoter activity using the green fluorescent protein (GFP) open reading frame (ORF) as a reporter (Fig. 1b). Whereas

this promoter activity was noticeably enhanced during meiosis, transcripts encoding GFP could be detected clearly in growing cells (Fig. 1c). This indicated that the *mei4* promoter was not completely downregulated during mitosis. We then examined the expression of *mei4* mRNA driven by the constitutive *adh1* promoter (Fig. 1d, top) and found that diploid cells carrying the *adh1-me4* chimaeric gene did not have detectable levels of *mei4* transcripts during mitotic growth (Fig. 1e, lane 1). If these diploid cells entered meiosis under nitrogen starvation, the *mei4* transcripts accumulated substantially (Fig. 1e, lanes 2, 3). The endogenous *adh1* mRNA was detected equally in mitotic and meiotic cells (data not shown). These observations indicated that an intrinsic feature of *mei4* mRNA prevented its accumulation in dividing yeast cells. The accumulation of *adh1-me4* transcripts could be induced in haploid cells if we artificially expressed a non-phosphorylatable form of Mei2 (Mei2SATA), which can provoke ectopic meiosis<sup>10</sup> (Fig. 1f, lane 2).

In subsequent analysis, deletion of a region 343 base pairs (bp) long within the *mei4* ORF, removing nucleotides (nt) 486 through 828 (Fig. 1d, bottom), caused the accumulation of the resulting mRNA in mitotic cells (Fig. 1f, lane 3). We hereafter refer to this deleted region as the 'determinant of selective removal' (DSR). We found that the transcripts of three other meiosis-specific genes, *ssm4*, *rec8* and *spo5*, were also selectively removed from mitotic cells. The *ssm4* gene encodes a homologue of the dynactin component Glued<sup>14,15</sup>, *rec8* encodes a meiosis-specific cohesin component<sup>16,17</sup>, and *spo5* (SPBC29A10.02/SPBC365.18) encodes an RNA recognition motif (RRM)-type RNA-binding protein required for meiosis II (our unpublished observations). By deletion analysis, we delimited their respective DSRs: *ssm4*DSR as 1,227–1,772 nt; *rec8*DSR as 1,781–2,029 nt and *spo5*DSR as 1,763–1,919 nt (Fig. 2a), although these segments do not necessarily represent the minimum essential regions. Whereas *ssm4*DSR was mapped within the ORF of the gene, as was *mei4*DSR, both the *rec8*DSR and *spo5*DSR sites span the border between the ORF and the 3'-UTR (Fig. 2a).

We constructed chimaeric genes carrying the *adh1* promoter, the

<sup>1</sup>Department of Biophysics and Biochemistry, Graduate School of Science, and <sup>2</sup>Molecular Genetics Research Laboratory, University of Tokyo, Hongo, Tokyo 113-0033, Japan.

<sup>3</sup>Kansai Advanced Research Center, National Institute of Information and Communications Technology, Kobe 651-2492, Japan. †Present addresses: Toukoudai Research Center, Astellas Pharma Inc., Toukoudai, Tsukuba 300-2698, Japan (H.T.); Institute of Molecular and Cellular Biosciences, University of Tokyo, Yayoi, Tokyo 113-0032, Japan (Y.W.).

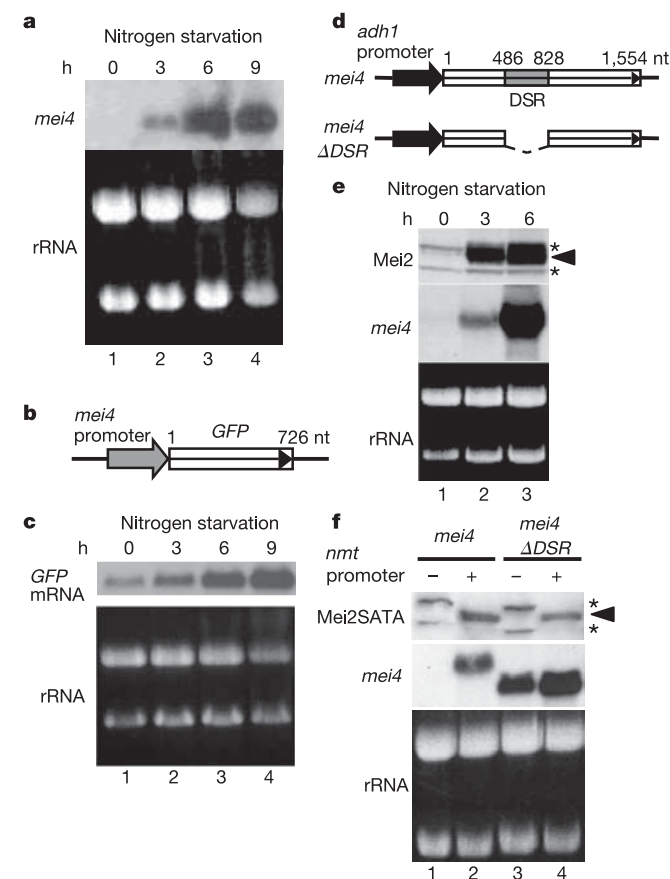
GFP ORF and one of the four DSRs (Fig. 2b). Expression of the control gene with no DSR was comparable between mitotic and Mei2SATA-induced meiotic cells (Fig. 2c, lanes 1, 2). The GFP–DSR transcripts were as abundant as the control in meiotic cells but scarce in mitotic cells (Fig. 2c, lanes 3–10), indicating that the presence of a DSR was sufficient to confer mitotic removal. These results suggested that fission yeast may possess a molecular machinery that regulates the selective elimination of meiotic mRNAs through their DSR sequences.

### Mmi1 controls elimination of mRNA

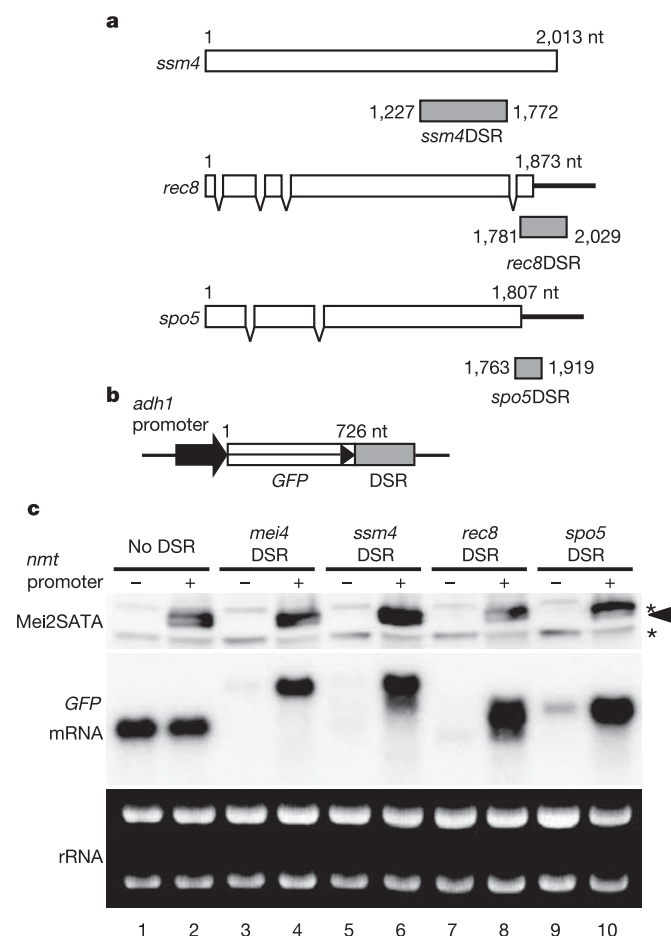
We screened for mutants that might be deficient in DSR-mediated

control of mRNA removal using a chimaeric *ura4* gene (Methods). Four candidates thus isolated expressed endogenous *mei4* and *ssm4* in the growing state (see below). Genetic crosses indicated that their mutations were recessive and constituted a single complementation group. We hence designated the mutated gene as *mmi1* (meiotic mRNA interception). To facilitate the cloning of *mmi1*, we created conditions in which the *mmi1* mutant was basically lethal but could survive if complemented by the wild-type gene (Methods). We subsequently obtained 11 candidate clones all carrying SPCC736.12c, encoding a putative RNA-binding protein of the YTH-family<sup>18</sup> (Fig. 3a). The identity of *mmi1* as SPCC736.12c was verified by sequencing the corresponding ORF in the four *mmi1* mutants. These all contained different point mutations in the carboxy-terminal region of SPCC736.12c (Fig. 3a). Reconstruction of these mutant alleles resulted in the *mmi1* phenotype, proving unequivocally that *mmi1* was SPCC736.12c.

Disruption of *mmi1* impairs cell growth severely (Fig. 3b), prompting us to construct temperature-sensitive *mmi1* alleles. We obtained two *ts* mutants, *mmi1-ts3* and *mmi1-ts6* (Fig. 3c). The



**Figure 1 | Selective removal of *mei4* mRNA in mitotic cells.** **a**, Expression profile of endogenous *mei4* mRNA during meiosis. Wild-type diploid cells (JY362) grown to mid-exponential phase were subjected to nitrogen starvation, and *mei4* mRNA was detected by northern blotting every 3 hours. The loading controls are rRNAs stained with ethidium bromide. **b**, Schematic illustration of the reporter construct to assay *mei4* promoter activity. **c**, Expression profile of the reporter construct. Diploid cells (JW740), in which the genomic *mei4* ORF was replaced with the GFP ORF, were analysed as in **a**, using the GFP sequence as a probe. **d**, Schematic illustration of the *mei4* constructs analysed in **e** and **f**. **e**, Expression profile of *mei4* mRNA driven from the constitutive *adh1* promoter. Diploid cells carrying the *adh1-me4* construct (JW772) were subjected to nitrogen starvation and analysed by northern blotting. The amount of Mei2 (arrowhead) was measured by western blotting using anti-Mei2 antibodies. Asterisks indicate non-specific signals. **f**, Expression of intact or deleted *mei4* mRNA in mitotic and Mei2SATA-induced meiotic cells. Haploid cells carrying either the *adh1-me4* construct (JW721; lanes 1, 2) or the *adh1-me4ΔDSR* (JW726; lanes 3, 4) besides the *nmt41-me2SATA* construct were incubated without thiamine for 16 hours to allow expression of Mei2SATA. Expression of *mei4* mRNA was measured by northern blotting before (lanes 1, 3) and after (lanes 2, 4) the Mei2SATA expression, which was confirmed by anti-Mei2 antibodies (arrowhead). Asterisks indicate non-specific signals.



**Figure 2 | DSRs found on meiosis-specific mRNAs.** **a**, Schematic illustration of DSRs assigned on *ssm4*, *rec8* and *spo5* mRNAs. *mei4*DSR is shown in Fig. 1d. **b**, A schematic of the constructs used to measure the ability of DSR to remove mRNA in growing cells. The GFP ORF was connected to each DSR and placed under control of the *adh1* promoter. **c**, Expression of GFP mRNA was examined in JX383 (*nmt41-me2SATA*) cells harbouring either pAGF-*mei4*DSR (lanes 3, 4), pAGT1-*ssm4*DSR (lanes 5, 6), pAGT1-*rec8*DSR (lanes 7, 8), or pAGF-*spo5*DSR (lanes 9, 10). The control plasmid pAGT1 bearing no DSR was examined in lanes 1 and 2. Odd-numbered lanes represent cells growing mitotically, whereas even-numbered lanes represent cells undergoing Mei2SATA-induced meiosis. pAGF differed from pAGT1 at the termination site and gave relatively slow-moving GFP transcripts.

expression of *mei4*, *ssm4*, *rec8* and *spo5* mRNAs was induced in these mutants within one hour after a temperature shift in nutrient medium (Fig. 3d). This confirmed that a common mechanism via *mmi1* governed the elimination of these DSR-containing mRNAs. Compared to the *ts* mutants, the expression of *rec8* and *spo5* mRNA was less distinct in the hypomorphic mutants isolated from the original screening (*mmi1*-48 and -619) (Fig. 3d, lanes 14, 15), suggesting that the original mutants contained a compromised

*mmi1* allele that permitted leaky expression of DSR-containing mRNAs and substantial cell growth. Importantly also, the growth defect of the *mmi1*-*ts* mutants at the restrictive temperature was alleviated by deletion of the *mei4* gene (Fig. 3c), suggesting that their growth retardation was caused mainly by inappropriate expression of *mei4* and genes activated by its product. Loss of *mmi1* function did not provoke meiosis by itself, although the cells exhibited aberrant traits that might be related to abortive meiosis (data not shown). Thus, this effect was phenotypically less conspicuous than the inactivation of Pat1 kinase or the activation of Mei2, which commits cells to unconditional meiosis<sup>10,19–21</sup>.

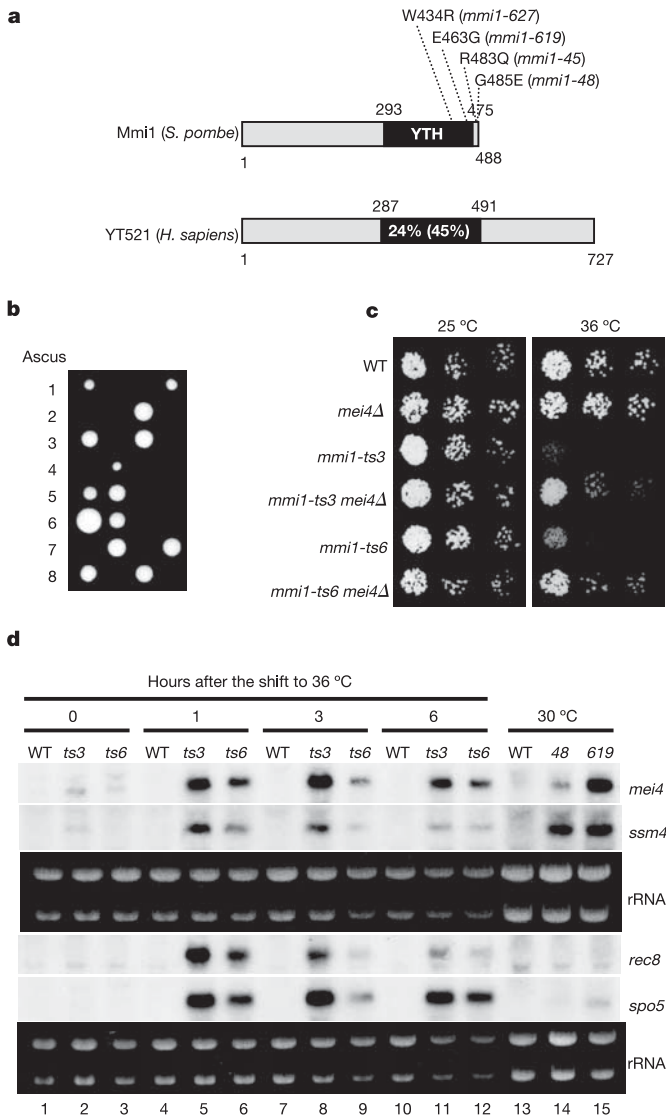
The overexpression of *mmi1* produced no significant influence on either mitotic cell growth or conjugation. However, cells overexpressing *mmi1* often failed to initiate pre-meiotic DNA synthesis, when examined by flow cytometry, and they sporulated poorly, indicating that Mmi1 can repress meiosis (Supplementary Fig. S2).

Because Mmi1 was a putative RNA-binding protein of the YTH family, we tested whether it could bind to DSR RNA, using electrophoretic mobility shift assays. This proved that Mmi1 could bind specifically to RNA molecules corresponding to the DSRs of the *mei4*, *ssm4*, *spo5* and *rec8* mRNAs (Supplementary Fig. S3). Mmi1 also revealed a moderate affinity for meiRNA, which binds to Mei2, but our analysis demonstrated no significant affinity between Mei2 and DSR RNA (Supplementary Fig. S3).

Hundreds of genes are transcriptionally upregulated during the initiation of meiosis in fission yeast<sup>3,22</sup>. We suspected that a considerable portion of them may be regulated by the *mmi1*-dependent mechanism and examined this possibility by analysing genome-wide profiles of gene expression in the *mmi1*-*ts3* mutant. By microarray assay, we found that 12 genes were upregulated more than twofold at the restrictive temperature, satisfying criteria stated in the Supplementary Methods (Table 1). Notable among these were *rec8*, *spo5* and *ssm4*, although *mei4* was not identified—its expression levels were judged to be below the detection limit of this experiment. This suggests that our microarray analysis was sufficiently selective but that some false negatives may exist. The 12 putative *mmi1* target genes in Table 1 include as many as nine early meiotic genes, two middle meiotic genes, and one unassigned, according to the classification in ref. 3. These results suggest that a large number of meiotic transcripts are susceptible to elimination by the DSR–*mmi1* system, so that their untimely functioning during mitosis is suppressed.

Mei2 sequesters Mmi1 during meiosis

Western blot analysis using anti-Mmi1 antibodies indicated that the levels of Mmi1 did not change significantly upon entry into meiosis (data not shown). However, its subcellular localization was altered substantially. In mitotic cells, Mmi1 was observed as one to several



**Figure 3 | The *mmi1* gene and expression of meiotic mRNAs in the absence of *mmi1* function.** **a**, A diagram of the structure of the *mmi1* gene product, which consists of 488 amino acids. The black box represents the region where human YT521 shows 24% identity and 45% similarity to Mmi1, which includes the YTH domain featuring in a family of RNA-binding proteins<sup>18</sup>. Single mutations found in the four hypomorphic *mmi1* alleles are indicated. **b**, Tetrad analysis of a *mmi1*<sup>+</sup>/*mmi1*::*kan*<sup>r</sup> diploid strain indicating that *mmi1* is indispensable for robust cell growth. **c**, Growth profiles of *mmi1*-*ts* mutants and *mmi1*-*ts* *mei4* double mutants at 25 °C and 36 °C. **d**, Expression of endogenous meiotic mRNAs in *mmi1* mutants. Cells of the wild-type (WT) (JV558), the *mmi1*-*ts3* (JV564), and the *mmi1*-*ts6* (JV567) strains were grown at 25 °C and shifted to 36 °C. Cells were sampled at the indicated time points and expression of DSR-containing meiotic mRNAs was examined by northern blotting (lanes 1–12). Expression of meiotic mRNAs was also examined in two hypomorphic *mmi1* mutants (*mmi1*-48; JV479 and *mmi1*-619; JV473) and the wild-type (JV761) grown at 30 °C (lanes 13–15). Loading controls are rRNAs stained with ethidium bromide.

**Table 1 | Genes regulated by *mmi1***

| ORF ID                   | Gene name                 | Class      | log <sub>2</sub> (Fold change) |
|--------------------------|---------------------------|------------|--------------------------------|
| SPAC6B12.16              | <i>meu26</i>              | Middle     | 2.75 ± 0.46                    |
| SPBC2G2.09c              | <i>crs1</i>               | Early      | 2.70 ± 0.40                    |
| SPCC70.09c               | ND                        | Early      | 2.53 ± 0.12                    |
| SPAC1556.06              | <i>meu1</i> , <i>meu2</i> | Middle     | 2.45 ± 0.13                    |
| SPBC29A10.14             | <i>rec8</i>               | Early      | 2.20 ± 0.01                    |
| SPAC6G9.13c              | <i>bqt1</i>               | Early      | 1.67 ± 0.05                    |
| SPAC458.04c              | ND                        | Early      | 1.53 ± 1.26                    |
| SPCC11E10.03             | ND                        | Early      | 1.43 ± 0.10                    |
| SPAC27D7.13c/SPAC637.01c | <i>ssm4</i>               | Early      | 1.31 ± 0.81                    |
| SPBC29A10.02/SPBC365.18  | <i>spo5</i>               | Early      | 1.23 ± 0.01                    |
| SPAC17A5.18c             | ND                        | Early      | 1.10 ± 0.56                    |
| SPCP20C8.03              | ND                        | Unassigned | 1.09 ± 0.09                    |

Twelve genes that satisfied the three criteria described in Supplementary Methods are shown. They are classified according to the Sanger Centre's database on the timing of meiotic gene expression (<http://www.sanger.uk>; ref. 3). The log<sub>2</sub>(fold change) compares the amounts of transcripts of each gene in the absence versus presence of the *mmi1* function, that is, after versus before the temperature shift. The value presented is an average of two independent sets of experiments and shown with the standard deviation (s.d.).



scattered nuclear foci (Fig. 4a). In meiotic cells, these foci converged into a single spot, reminiscent of the nuclear Mei2 dot. Co-expression of cyan fluorescent protein (CFP)-tagged Mmi1 and yellow fluorescent protein (YFP)-tagged Mei2 indicated that CFP-Mmi1 was indeed localized to the Mei2p dot (Fig. 4b). The site for Mmi1 convergence overlapped with the Mei2 dot in all of the 74 cells examined. Furthermore, this convergence did not occur in the *mei2Δ* or *sme2Δ* (that is, meiRNA-missing) strains (Fig. 4c), which fail to develop the dot<sup>11</sup>. The behaviour of Mmi1 during meiosis was analysed in cells expressing GFP-Mmi1 and CFP-tagged  $\alpha$ -tubulin (Atb2)<sup>23</sup>; CFP-Atb2 serves as an indicator of the meiotic stage. Mmi1

converged into a dot in conjugating cells, persisted as a dot during the horsetail-shaped nucleus (meiotic prophase) period and dispersed into multiple nuclear foci after meiosis I (Fig. 4d). Thus, the duration of the converged Mmi1 dot paralleled that of the Mei2 dot described previously<sup>11</sup>. These results support the contention that the Mei2 dot anchors Mmi1.

We examined the possible interaction of Mmi1 with Mei2 by GST pull-down assays. As shown in Fig. 4e, GST-Mmi1 specifically pulled down *in vitro*-translated Mei2 from a rabbit reticulocyte lysate, indicating direct binding. The C-terminal half of Mei2 appeared to have higher affinity for Mmi1 than the amino-terminal half (Fig. 4e).

Forced localization of Mei2 to the nucleus by use of a nuclear localizing signal (NLS) can relieve the meiotic arrest of *sme2Δ* cells<sup>11</sup>. Mei2-NLS forms a dot-like structure in the nucleus if expressed in mitotic cells<sup>24</sup>. The proportion of mitotic cells carrying a single Mmi1 focus was about 25%, but if Mei2-NLS was expressed, this value rose to 60% and the cells exhibited DSR-containing transcripts in the growing state (Supplementary Fig. S4a–c). This indicates that Mei2 localized to the nucleus can boost meiosis-specific transcripts.

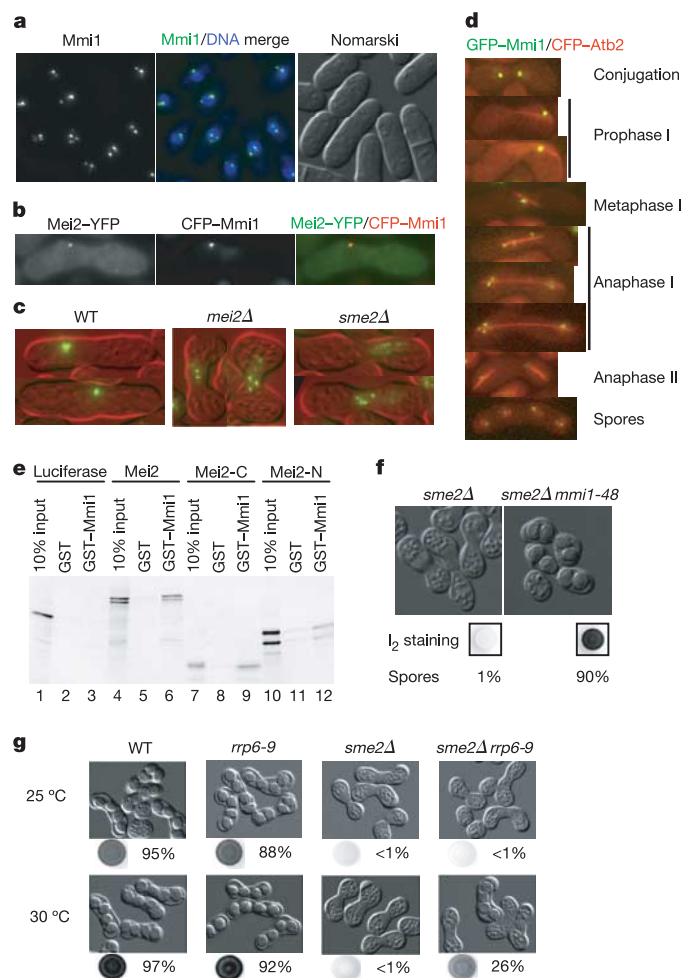
All of these observations led us to hypothesize that the Mei2 dot might sequester Mmi1 during early phases of meiosis, so that meiosis-specific mRNAs are not intercepted by Mmi1 and can achieve stable expression. To test this idea, we examined whether the arrest of meiosis due to the absence of the Mei2 dot could be suppressed by a decrease in Mmi1 activity. As shown in Fig. 4f, *mmi1-48*, one of the hypomorphic *mmi1* mutations, could suppress loss of meiRNA. The *sme2Δ mmi1-48* strain underwent meiosis and generated four-spored asci at a frequency of 90%, whereas the control *sme2Δ* strain hardly sporulated (1%).

We had noticed for some time that the overexpression of DSR RNA could suppress *sme2Δ* (ref. 14 and Supplementary Fig. S4d). One possible explanation for this might be that DSR RNA substituted for meiRNA and promoted the function of Mei2. However, DSR RNA does not substantially bind Mei2, as shown above, and the most likely explanation is therefore that excess DSR RNA sequesters Mmi1 and consequently allows expression of meiosis-specific mRNAs. This strengthens the idea that the critical function of the Mei2 dot is to sequester Mmi1.

### Molecular mechanisms and implications

Mmi1 and the mammalian YTH-family protein YT521-B<sup>25</sup> share only limited homology but apparently have similar properties. YT521-B localizes to multiple nuclear foci, known as YT bodies, which contain transcriptionally active sites<sup>26</sup>. It is proposed that tyrosine phosphorylation causes sequestration of YT521-B and converts it to an insoluble nuclear form, eventually abolishing its ability to change alternative splice site selection<sup>27</sup>. Thus, regulation through relocation may be a common mechanism adopted by the YTH-family proteins. It would be interesting to test the involvement of YTH-family proteins in gene regulation during gametogenesis or in general gene silencing in higher eukaryotes.

Hereafter we consider possible mechanisms underlying the Mmi1-mediated elimination of meiotic mRNAs. DSRs from *mei4*, *ssm4*, *rec8* and *spo5* have been delimited into a few hundred nucleotides but they show no obvious similarity in either the primary or the secondary structure. Therefore we suspect that the DSR function may be the sum of small contributions from multiple and redundant elements within it. Our RNA polymerase II-chromatin immunoprecipitation experiments excluded the possibility that the presence of a DSR on an mRNA somehow downregulates transcription of the gene (Supplementary Fig. S5a). We then measured half-lives of DSR-containing mRNAs in mitotic and meiotic cells (Supplementary Fig. S5b). The half-life of *mei4* mRNA in growing cells was only 2.8 min and the presence of a DSR appeared to shorten the mitotic half-life up to twofold, paralleling an early observation in *S. cerevisiae*



**Figure 4 | The Mei2 dot anchors Mmi1.** **a**, Live observation of GFP-Mmi1 in vegetatively growing cells (JV986). Nuclei are counterstained with Hoechst 33342. **b**, Live images of a meiotic prophase cell expressing CFP-Mmi1 and Mei2-YFP (JV866). **c**, Scattered GFP-Mmi1 in zygotes defective in *mei2* (JV872) or *sme2* (JV873). Wild-type: JV871. **d**, Live observation of GFP-Mmi1 in JV871 cells undergoing zygotical meiosis. The meiotic stage of each cell was assigned from the microtubule organization visualized by CFP-Atb2 expressed from a plasmid. **e**, Interaction between Mei2 and Mmi1. *In vitro* translated full-length Mei2 or its C-terminal (amino acids 429–750) or N-terminal (1–429) half, labelled with <sup>35</sup>S-methionine, was incubated with either GST-Mmi1 or GST, and subjected to GST pull-down assays. Binding was assessed by SDS-polyacrylamide gel electrophoresis (PAGE) followed by autoradiography. Luciferase was used as a negative control (lanes 1–3). **f**, Suppression of the meiotic arrest in *sme2Δ* cells by the *mmi1-48* mutation. Sporulation of JZ464 (*sme2Δ*) and JV859 (*sme2Δ mmi1-48*) was analysed either by exposure of a patch culture to I<sub>2</sub> vapour, which stains spores dark brown, or by counting ascospores under a microscope. **g**, Suppression of *sme2Δ* by a temperature-sensitive *rrp6-9* mutation. I<sub>2</sub>-staining and sporulation efficiency were measured as in **f**. Construction of the *rrp6-9* allele will be described elsewhere.

that some meiotic mRNAs likewise show short half-lives in mitotic cells<sup>28</sup>. However, a twofold difference is insufficient to explain the intense removal of meiotic mRNAs from mitotic cells, suggesting that cytoplasmic mature mRNA is unlikely to be the main target of elimination.

The above observations and the exclusive localization of Mmi1 to the nucleus (Fig. 4a) strongly suggest that the elimination of meiotic mRNA is an early event that follows transcription. Mmi1, which has no significant homology with nucleases, may cooperate with an mRNA destruction system. Our analyses indicated that none of exonuclease II<sup>29</sup>, RNase III (pac1 endonuclease)<sup>30</sup>, the RNAi pathway<sup>31,32</sup> or the nonsense-mediated decay system<sup>33</sup> was relevant to the elimination (unpublished observations), but a temperature-sensitive mutation in *rrp6*, the orthologue of *S. cerevisiae* RRP6 encoding a nucleus-specific subunit of exosome<sup>34</sup>, could suppress *sme2Δ* (Fig. 4g). Furthermore, Mmi1 and Rrp6 interacted physically (unpublished observations). *S. cerevisiae* Rrp6 is required for 3' end processing of ribosomal RNAs, small nuclear RNAs, and small nucleolar RNAs<sup>35</sup>, and its role in mRNA surveillance is also suggested<sup>36</sup>. Therefore a plausible hypothesis emerges that Mmi1 bound to a DSR may activate nuclear exosome, which may then degrade the DSR-containing transcript from the 3' end. More rigorous evidence is awaited to validate this hypothesis.

We wonder why the control of meiosis is so complex. All eukaryotic cells possess meiotic potential but most never initiate it. It would be deleterious if meiosis-specific genes were unconditionally expressed in cells undergoing mitosis, but blocking the initiation of transcription may not be sufficient to prevent this. Hence, a fail-safe mechanism against unnecessary meiotic gene expression may have been developed, which would need to be deactivated at the onset of meiosis, as is performed by Mei2 in fission yeast. DSRs embedded in the ORF or the 3'-UTR of meiosis-specific transcripts appear to demonstrate another tier of biological regulation. The elegance of this system, which has enabled each meiotic gene to incorporate a respective DSR sequence into mRNA without damaging the function of its protein product, is remarkable.

## METHODS

**Fission yeast strains, plasmids, genomic library and antibodies.** *S. pombe* genotypes are listed in Supplementary Table S1. Vectors carrying the thiamine-repressible *nmt1* promoter or its derivative have been described previously<sup>37,38</sup>, as have the *S. pombe* genomic library and anti-Mei2 antibodies<sup>9</sup>.

**Mutants, cloning, disruption and mutagenesis of the *mmi1* gene.** The original hypomorphic *mmi1* mutants were isolated using a chimaeric *ura4* gene, which carried the *mei4*DSR near the 3' end of the *ura4* ORF. Mutants that could stably express this *ura4-me4*DSR mRNA were selected, as detailed in Supplementary Methods. To clone the *mmi1* gene, we established conditions in which the *mmi1* mutation, and hence stable expression of a DSR-containing mRNA, would cause cell death, using the fatal *rec8RDRD* allele<sup>39</sup>. Further details of cloning are given in Supplementary Methods. Analysis of *mmi1* (SPCC736.12c) complementary DNA revealed the presence of the fourth intron from 1,539 nt through to 1,582 nt not assigned in the Sanger Centre database. Reassignment of the termination codon established that the *mmi1* gene encodes 488 amino acids. Gene disruption of *mmi1* and construction of the temperature-sensitive *mmi1-ts3* and *mmi1-ts6* alleles are also described in Supplementary Methods.

**Microarray analysis.** We searched for genes whose expression was enhanced in the *mmi1-ts3* mutant at the restrictive temperature, using a fission yeast cDNA microarray that covered more than 4,900 genes, that is, most of the protein-coding genes in this microbe. Preparation of cDNA microarrays was done essentially as described<sup>40</sup>, and general protocols of RNA isolation from fission yeast, sample labelling, microarray hybridization and data processing will be described elsewhere (Y.C. and Y. Hiraoka, manuscript in preparation). Specific conditions employed in the current analysis and the criteria used to select positives are given in Supplementary Methods.

Received 26 January; accepted 5 May 2006.

- Chu, S. *et al.* The transcriptional program of sporulation in budding yeast. *Science* **282**, 699–705 (1998).
- Primig, M. *et al.* The core meiotic transcriptome in budding yeasts. *Nature Genet.* **26**, 415–423 (2000).
- Mata, J., Lyne, R., Burns, G. & Bahler, J. The transcriptional program of meiosis and sporulation in fission yeast. *Nature Genet.* **32**, 143–147 (2002).
- Kupeic, M., Byers, B., Esposito, R. E. & Mitchell, A. P. in *The Molecular and Cellular Biology of the Yeast Saccharomyces* (eds Pringle, J., Broach, J. & Jones, E.) 889–1036 (Cold Spring Harbor Laboratory Press, New York, 1997).
- Yamamoto, M., Imai, Y. & Watanabe, Y. in *The Molecular and Cellular Biology of the Yeast Saccharomyces* (eds Pringle, J., Broach, J. & Jones, E.) 889–1036 (Cold Spring Harbor Laboratory Press, New York, 1997).
- Kunitomo, H., Higuchi, T., Iino, Y. & Yamamoto, M. A zinc-finger protein, Rst2p, regulates transcription of the fission yeast *ste11<sup>+</sup>* gene, which encodes a pivotal transcription factor for sexual development. *Mol. Biol. Cell* **11**, 3205–3217 (2000).
- Higuchi, T., Watanabe, Y. & Yamamoto, M. Protein kinase A regulates sexual development and gluconeogenesis through phosphorylation of the Zn finger transcriptional activator Rst2p in fission yeast. *Mol. Cell. Biol.* **22**, 1–11 (2002).
- Sugimoto, A., Iino, Y., Maeda, T., Watanabe, Y. & Yamamoto, M. *Schizosaccharomyces pombe ste11<sup>+</sup>* encodes a transcription factor with an HMG motif that is a critical regulator of sexual development. *Genes Dev.* **5**, 1990–1999 (1991).
- Watanabe, Y. & Yamamoto, M. *S. pombe mei2<sup>+</sup>* encodes an RNA-binding protein essential for premeiotic DNA synthesis and meiosis I, which cooperates with a novel RNA species *meiRNA*. *Cell* **78**, 487–498 (1994).
- Watanabe, Y., Shinozaki-Yabana, S., Chikashige, Y., Hiraoka, Y. & Yamamoto, M. Phosphorylation of RNA-binding protein controls cell cycle switch from mitotic to meiotic in fission yeast. *Nature* **386**, 187–190 (1997).
- Yamashita, A., Watanabe, Y., Nukina, N. & Yamamoto, M. RNA-assisted nuclear transport of the meiotic regulator Mei2p in fission yeast. *Cell* **95**, 115–123 (1998).
- Shimada, T., Yamashita, A. & Yamamoto, M. The fission yeast meiotic regulator Mei2p forms a dot structure in the horse-tail nucleus in association with the *sme2* locus on chromosome II. *Mol. Biol. Cell* **14**, 2461–2469 (2003).
- Horie, S. *et al.* The *Schizosaccharomyces pombe mei4<sup>+</sup>* gene encodes a meiosis-specific transcription factor containing a forkhead DNA-binding domain. *Mol. Cell. Biol.* **18**, 2118–2129 (1998).
- Yamashita, A., Watanabe, Y. & Yamamoto, M. Microtubule-associated coiled-coil protein Ssm4 is involved in the meiotic development in fission yeast. *Genes Cells* **2**, 155–166 (1997).
- Niccoli, T., Yamashita, A., Nurse, P. & Yamamoto, M. The p150-Glued Ssm4p regulates microtubular dynamics and nuclear movement in fission yeast. *J. Cell Sci.* **117**, 5543–5556 (2004).
- Watanabe, Y. & Nurse, P. Cohesin Rec8 is required for reductional chromosome segregation at meiosis. *Nature* **400**, 461–464 (1999).
- Parisi, S. *et al.* Rec8p, a meiotic recombination and sister chromatid cohesion phosphoprotein of the Rad21p family conserved from fission yeast to humans. *Mol. Cell. Biol.* **19**, 3515–3528 (1999).
- Stoilov, P., Rafalska, I. & Stamm, S. YTH: a new domain in nuclear proteins. *Trends Biochem. Sci.* **27**, 495–497 (2002).
- Iino, Y. & Yamamoto, M. Mutants of *Schizosaccharomyces pombe* which sporulate in the haploid state. *Mol. Gen. Genet.* **198**, 416–421 (1985).
- Nurse, P. Mutants of the *Schizosaccharomyces pombe* which alter the shift between cell proliferation and sporulation. *Mol. Gen. Genet.* **198**, 497–502 (1985).
- McLeod, M. & Beach, D. A specific inhibitor of the *ran1<sup>+</sup>* protein kinase regulates entry into meiosis in *Schizosaccharomyces pombe*. *Nature* **332**, 509–514 (1988).
- Watanabe, T. *et al.* Comprehensive isolation of meiosis-specific genes identifies novel proteins and unusual non-coding transcripts in *Schizosaccharomyces pombe*. *Nucleic Acids Res.* **29**, 2327–2337 (2001).
- Glynn, J. M., Lustig, R. J., Berlin, A. & Chang, F. Role of bud6p and tea1p in the interaction between actin and microtubules for the establishment of cell polarity in fission yeast. *Curr. Biol.* **11**, 836–845 (2001).
- Sato, M., Shinozaki-Yabana, S., Yamashita, A., Watanabe, Y. & Yamamoto, M. The fission yeast meiotic regulator Mei2p undergoes nucleocytoplasmic shuttling. *FEBS Lett.* **499**, 251–255 (2001).
- Hartmann, A. M., Nayler, O., Schwaiger, F. W., Obermeier, A. & Stamm, S. The interaction and colocalization of Sam68 with the splicing-associated factor YT521-B in nuclear dots is regulated by the Src family kinase p59<sup>lyn</sup>. *Mol. Biol. Cell* **11**, 3909–3926 (1999).
- Nayler, O., Hartmann, A. M. & Stamm, S. The ER repeat protein YT521-B localizes to a novel subnuclear compartment. *J. Cell Biol.* **150**, 949–962 (2000).
- Rafalska, I. *et al.* The intranuclear localization and function of YT521-B is regulated by tyrosine phosphorylation. *Hum. Mol. Genet.* **13**, 1535–1549 (2004).
- Surosky, R. T. & Esposito, R. E. Early meiotic transcripts are highly unstable in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **12**, 3948–3958 (1992).
- Szankasi, P. & Smith, G. R. Requirement of *S. pombe* exonuclease II, a homologue of *S. cerevisiae* Sep1, for normal mitotic growth and viability. *Curr. Genet.* **30**, 284–293 (1996).
- Iino, Y., Sugimoto, A. & Yamamoto, M. *S. pombe pac1<sup>+</sup>*, whose overexpression inhibits sexual development, encodes a ribonuclease III-like Rnase. *EMBO J.* **10**, 221–226 (1991).

31. Hansen, K. R. *et al.* Global effects on gene expression in fission yeast by silencing and RNA interference machineries. *Mol. Cell. Biol.* **25**, 590–601 (2005).
32. Cam, H. P. *et al.* Comprehensive analysis of heterochromatin- and RNAi-mediated epigenetic control of the fission yeast genome. *Nature Genet.* **37**, 809–819 (2005).
33. Mendell, J. T. *et al.* Novel Upf2p orthologues suggest a functional link between translation initiation and nonsense surveillance complexes. *Mol. Cell. Biol.* **20**, 8944–8957 (2000).
34. Allmang, C. *et al.* The yeast exosome and human PM-Scl are related complexes of 3'→5' exonucleases. *Genes Dev.* **13**, 2148–2158 (1999).
35. Jensen, T. H. & Moore, C. Reviving the exosome. *Cell* **121**, 660–662 (2005).
36. Milligan, L., Torchet, C., Allmang, C., Shipman, T. & Tollervey, D. A nuclear surveillance pathway for mRNAs with defective polyadenylation. *Mol. Cell. Biol.* **25**, 9996–10004 (2005).
37. Maundrell, K. Thiamine-repressible expression vectors pREP and pRIP for fission yeast. *Gene* **123**, 127–130 (1993).
38. Basi, G., Schmid, E. & Maundrell, K. TATA box mutations in the *Schizosaccharomyces pombe nmt1* promoter affect transcription efficiency but not the transcription start point or thiamine repressibility. *Gene* **123**, 131–136 (1993).
39. Kitajima, T. S., Miyazaki, Y., Yamamoto, M. & Watanabe, Y. Rec8 cleavage by separase is required for meiotic nuclear divisions in fission yeast. *EMBO J.* **22**, 5643–5653 (2003).
40. Zhao, S. & Bruce, W. B. Expression profiling using cDNA microarrays. *Methods Mol. Biol.* **236**, 365–380 (2003).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature). A figure summarizing the main results of the paper is also included.

**Acknowledgements** We thank A. M. Carr for providing the chromosomal N-terminal tagging system, G. R. Smith for fission yeast strains, and M. Sato for help in strain construction. This work was supported by a Grant-in-Aid for Specially Promoted Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan (to M.Y.) and a Grant from the Japan Science and Technology Agency (to Y. Hiraoka). Y. Harigaya was a recipient of a JSPS Research Fellowship for Young Scientists (DC) associating with a 21st Century COE Program led by M.Y.

**Author Information** The entire microarray data obtained in this study has been deposited in the Gene Expression Omnibus (GEO), under the accession number GSE3314. Reprints and permissions information is available at [npg.nature.com/reprintsandpermissions](http://npg.nature.com/reprintsandpermissions). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.Y. ([yamamoto@biochem.s.u-tokyo.ac.jp](mailto:yamamoto@biochem.s.u-tokyo.ac.jp)).



# Detection of Earth-like planets around nearby stars using a petal-shaped occulter

Webster Cash<sup>1</sup>

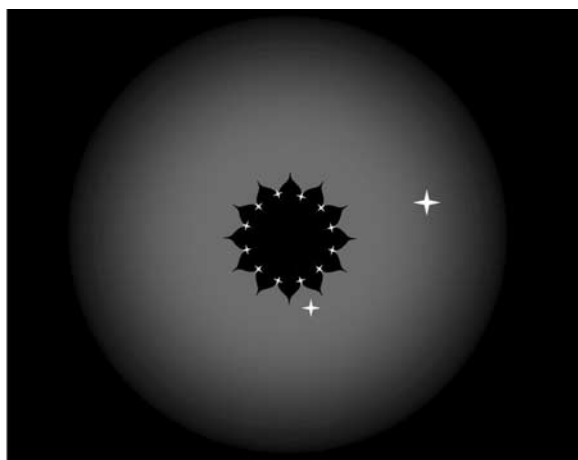
Direct observation of Earth-like planets is extremely challenging, because their parent stars are about  $10^{10}$  times brighter but lie just a fraction of an arcsecond away<sup>1</sup>. In space, the twinkle of the atmosphere that would smear out the light is gone, but the problems of light scatter and diffraction in telescopes remain. The two proposed solutions—a coronagraph internal to a telescope and nulling interferometry from formation-flying telescopes—both require exceedingly clean wavefront control in the optics<sup>2</sup>. An attractive variation to the coronagraph is to place an occulting shield outside the telescope, blocking the starlight before it even enters the optical path<sup>3</sup>. Diffraction and scatter around or through the occulter, however, have limited effective suppression in practically sized missions<sup>4–6</sup>. Here I report an occulter design that would achieve the required suppression and can be built with existing technology. The compact mission architecture of a coronagraph is traded for the inconvenience of two spacecraft, but the daunting optics challenges are replaced with a simple deployable sheet 30 to 50 m in diameter. When such an occulter is flown in formation with a telescope of at least one metre aperture, terrestrial planets could be seen and studied around stars to a distance of ten parsecs.

A starshade suitable for planet hunting must be designed such that the diffracted light is minimized. This means the sum of the phases of the light from all the paths through and around the shade must be extremely close to zero, implying a wide range of phases in the focal

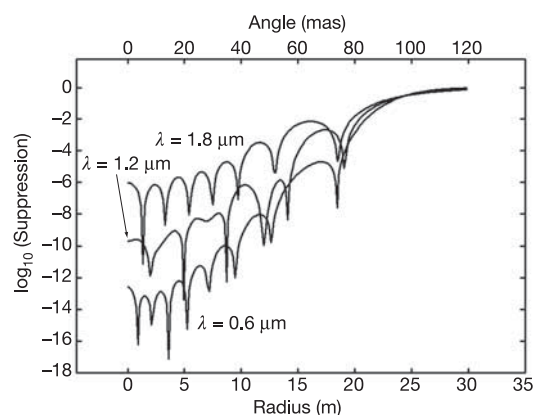
plane. The number of extra wavelengths ( $m\lambda$ ) a ray must travel to reach the centre of the shadow around an occulter of radius  $R$  at a distance  $F$  is given by  $R^2 = 2m\lambda F$ . Noting that  $R/F$  is the angular diameter of the shade, the relation becomes  $R = 2m\lambda/\theta$ . For planet-finding  $\theta$  is  $5 \times 10^{-7}$ , so a planet just 0.1 arcsec from its parent star may be detected, and in the visible band  $\lambda = 6 \times 10^{-7}$  m. If  $m$  is at least 10, then to create a large range of incident phases,  $R$  must be at least 20 m. So, based only on wavelength and planet–star angle, one finds that the starshade must be a large distance ( $R/\theta \approx 40,000$  km) from the telescope. Conveniently, occulters with diameters of tens of metres can also fully shade the large (up to 10 m in diameter) telescopes suitable for studying the planets.

A recent study of transmitting apertures showed it was possible to efficiently suppress diffraction over a broad spectral band to the  $10^{-10}$  level very close to a stellar image<sup>7</sup>. The results had the further key feature of being ‘binary’ (either fully transmitting or fully opaque at each and every point), thus avoiding the problem of a partially transmitting sheet that would be difficult to manufacture to the needed tolerances and might reintroduce scatter. Then, a class of circularly symmetric apertures was shown to enable suppression in all directions simultaneously between some inner and outer working angles<sup>8</sup>. These apertures could be made binary and still function well by approximating the circularly symmetric fall-off with an array of petal-shaped apertures.

By numerical integration of the Fresnel diffraction equations and by subsequent mathematical derivation (provided in the Supplementary Information) I have shown there exists a class of shaping



**Figure 1 | Schematic showing how a starshade in position against a nearby star might appear.** If a telescope of sufficiently good quality is stationed at the centre of the shadow, it would see the shade outlined against the zodiacal light of the target star system. At the base of each petal a small amount of light from the parent star diffracts around the shade. Planets simply appear as faint stars in the field of view. The shape of the starshade is based on the parameters used to calculate performance in Fig. 2.



**Figure 2 | Achieved stellar suppression.** The stellar suppression ratio of a typical starshade is shown for three wavelengths:  $\lambda = 0.6 \mu\text{m}$ ,  $1.2 \mu\text{m}$  and  $1.8 \mu\text{m}$ . The depth of the shadow is plotted versus the radius across the bottom of the graph and against the effective angle across the top. The transmission is nearly 100% just 0.1 arcsec off-axis, while the suppression ratio is below the desired  $10^{-10}$  for the two shorter wavelengths at the centre of the shadow. The design values are  $a = b = 12.5$  m,  $n = 6$ ,  $F = 50,000$  km.

<sup>1</sup>Center for Astrophysics and Space Astronomy, University of Colorado, Boulder, Colorado 80309, USA.

**Table 1 | Characteristics of the occulter compared to the coronagraph**

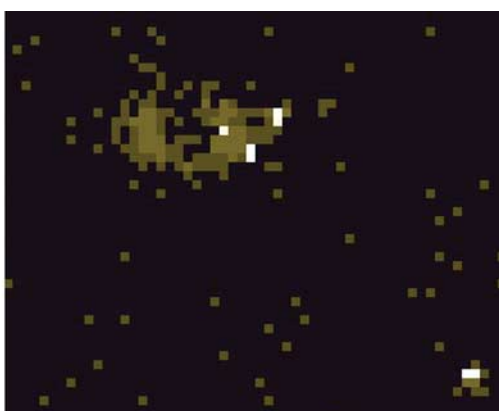
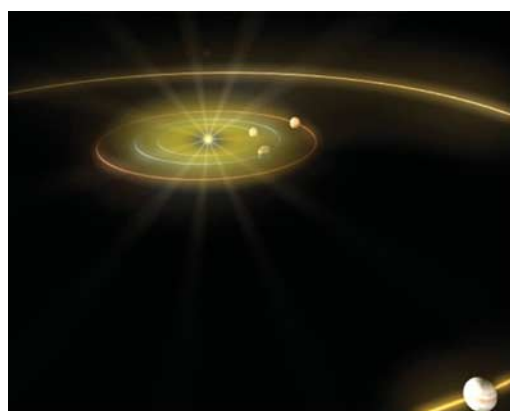
|                                     | Coronagraph |                   | Occulter |                        |
|-------------------------------------|-------------|-------------------|----------|------------------------|
| <b>Mission</b>                      |             |                   |          |                        |
| Architecture                        | +           | Single spacecraft | +        | Multiple spacecraft    |
| Telescope                           |             | Dedicated         |          | Existing               |
| Mechanical                          |             | Monolithic mirror |          | Large deployable sheet |
| Field of regard                     | +           | All sky           |          | Restricted             |
| <b>Optics</b>                       |             |                   |          |                        |
| Wavefront control                   |             | Challenging       | +        | None                   |
| Diffraction                         |             | Shaped aperture   | +        | Shaped occulter        |
| Scatter                             |             | Internal optics   | +        | Edge scatter           |
| Degradation in flight               |             | Risky             | +        | Robust                 |
| Technical readiness                 |             | Poor              | +        | Good                   |
| <b>Science</b>                      |             |                   |          |                        |
| Sensitivity to Earth-like planets   |             | Good              |          | Good                   |
| Sensitivity to Jupiter-like planets |             | Limited           | +        | Excellent              |
| Sensitivity to disks                |             | Poor              | +        | Excellent              |
| Targeting                           | +           | Fast              |          | Slow                   |

A range of mission considerations for the occulter and coronagraph approaches are compared. Features providing a significant advantage to one approach or the other are marked with a plus sign. Coronagraphs are single-spacecraft missions allowing quick, wide-field response, whereas occulters can take advantage of existing telescopes. The occulters are easier to implement in all areas of the optics. Scientifically, both approaches are sensitive to Earth-like planets, but the occulters provide extra sensitivity to outer planetary systems and debris disks. Overall, the large, high-quality optics push the cost of coronagraph missions significantly higher than the cost of occulter missions. I note that an occulter can be used with any existing or planned telescope (like JWST) whereas a coronagraph must be specially designed and built. A coronagraph can view (field of regard) most of the sky at any given time, whereas occulters are restricted to lines of sight perpendicular to the sun-spacecraft line to reduce scattered sunlight. All stars can be studied by either, but a coronagraph can revisit a system more often. Coronagraphs scatter starlight inside the optics, whereas occulters scatter sunlight off their edges—easier to control. Pitting from micrometeor impact or reflectivity loss from surface contamination (degradation in flight) can cause scatter in coronagraphs. Occulters can suffer tiny holes from micrometeor impacts, but a triple layer of material keeps the shade from transmitting starlight. Coronagraphs still need development of basic optical technology, while the technology for occulters already exists (technical readiness). The presence of an outer working angle keeps coronagraphs from simultaneously observing inner and outer planetary systems (planet sensitivity). Coronagraphs have high residual internal background, which makes extended sources like disks difficult to detect (debris sensitivity).

(apodization) functions that creates high levels of diffraction control for occulting masks. The diffraction suppression factors achieved are many orders of magnitude more effective than predecessor designs at any given size and finally allow the design of practically sized occulters. The function (shown in binary form in Fig. 1) is:

$$A(\rho) = 1 - e^{-(\rho-a)/b)^n} \text{ for } \rho > a$$

$$A(\rho) = 0 \text{ for } \rho < a$$



**Figure 3 | Seeing Earth-like planets.** The left panel is an artist's conception of our Solar System as viewed from outside. The planets from Venus to Jupiter are shown along with their orbits. The glow around the Sun represents the zodiacal light, caused by dust scattering in the inner Solar System. In the right panel I show a simulation of how this system would

This bipartite function features an opaque central circle of radius  $a$ , and petals that start at  $a$  and rise to  $1 - e^{-1}$  at a radius of  $a + b$ . The suppression ratio  $S$  of diffraction in the centre of the shadow, relative to the unobscured light, is given by:

$$S \leq (n!)^2 (\lambda F / 2\pi ab)^{2n}$$

in a well-designed occulter system large enough that  $b^2 \gg n(F\lambda/2\pi)$ . In the Supplementary Information I show that the solution is remarkably robust and scale invariant. I also estimate the range of the allowed fabrication errors and establish they are within the current state of the art of aerospace engineering.

In Fig. 1 I show the shape of an occulter where  $a = b = 12.5$  m,  $n = 6$ ,  $F = 50,000$  km as it might appear through a telescope in alignment with a nearby star. Figure 2 shows the suppression ratio to be expected with such an occulter at three wavelengths. The off-axis angle is plotted across the top of the graph, showing that a planet just 0.1 arcsec off-axis will reach the telescope essentially unobscured. For the shorter wavelengths, the deep zone of the shadow is 10–20 m in diameter, wide enough to hide a large telescope with room to spare.

There may eventually be other applications for optimally deep shadows, but the invention has been applied to the planet-finding mission named The New Worlds Observer<sup>9</sup>. It comprises two spacecraft, a flower-shaped starshade about 50 m from tip to tip, and a conventional-quality telescope. The telescope optic must be diffraction-limited in the visible band and at least 1 m in diameter. The mission operates by flying the starshade into the line of sight of a nearby star, a move that can take several days. After alignment, the starshade must hold its position against tidal forces by using small thrusters.

The New Worlds Observer offers major advantages over other approaches to direct studies of nearby planetary systems. In Table 1 I compare its strengths and weaknesses with the Terrestrial Planet Finder Coronagraph (TPF-C), which has been extensively studied<sup>2</sup>. The TPF-C is a single-spacecraft observatory that features a very large, very-high-quality telescope, correcting optics for wavefront control, and a pupil-plane coronagraph. It requires a 6–8-m-class monolithic mirror that corrects its optical figure to better than  $\lambda/5,000$  and holds or corrects its reflection uniformity to high precision across a broad spectral band after several years in orbit. From Table 1 it is clear that the challenges the New Worlds Observer must meet are the deployment of a large shade and automated station-keeping relative to the celestial sphere, but these attributes can be added to any space telescope already in deep space. Overall, the large, high-quality optics of TPF-C are more expensive and risky than the extra spacecraft with the large, deployable sheet needed for the

look when viewed from a distance of 7 parsec using an occulter and the James Webb Space Telescope. The strength of the zodiacal light has been reduced a factor of ten over solar and is still clearly visible, with the dark spot in the middle caused by the starshade itself. Earth, Mars and Jupiter are visible as bright white spots. (Artwork courtesy of Ball Aerospace.)

New Worlds Observer. Surprisingly, because occulter can observe over a wider angular range with lower noise, they are significantly more sensitive to outer planets (like Jupiter and Saturn) and to debris disks.

Simulations such as those in Fig. 3 show that the New Worlds Observer will be able to photograph directly the major features of planetary systems to ten parsecs and beyond. It can detect all the major planets (from the habitable zone outward), the zodiacal light, debris disks and possibly even comets. Photometric variations might show the presence of surface features like oceans and continents. Follow-up spectroscopy of the detected planets would enable classification by type, and the presence of water would be clearly visible in atmospheric absorption lines. Atmospheric markers (like free oxygen absorption lines) could potentially provide the first evidence of life outside our Solar System.

The New Worlds Observer is actively being studied in academia, industry and government. The technical challenges related to building such a mission appear to be within the reach of existing engineering and within the budgetary scope of NASA's science programme. One particularly appealing option is to fly a starshade to work in concert with the upcoming, powerful James Webb Space Telescope. If Earth-like planets around the nearby stars do exist, use of an occulter could find them within the next decade.

Received 16 January; accepted 19 May 2006.

1. Seager, S. The search for extra-solar Earth-like planets. *Earth Planet. Sci. Lett.* **208**, 113–124 (2003).
2. Henry, C. *et al.* Terrestrial planet finder interferometer: architecture, mission

design, and technology development. *Proc. Soc. Photo-Opt. Instrum. Eng.* **5491**, 265–274 (2004).

3. Marchal, C. Concept of a space telescope able to see the planets and even the satellites around the nearest stars. *Acta Astronaut.* **12**, 195–201 (1985).
4. Copi, C. J. & Starkman, G. D. The big occulting steerable satellite (BOSS). *Astrophys. J.* **532**, 581–592 (2000).
5. Schultz, A. B. *et al.* UMBRAS: a matched occulter and telescope for imaging extrasolar planets. *Proc. Soc. Photo-Opt. Instrum. Eng.* **4860**, 54–61 (2003).
6. Jordan, I. J. E. *et al.* Apodized square aperture plus occulter concept for TPF. *Proc. Soc. Photo-Opt. Instrum. Eng.* **5487**, 1391–1401 (2004).
7. Kasdin, N. J., Vanderbei, R. J., Spergel, D. N. & Littman, M. G. Extrasolar planet finding via optimal apodized and shaped pupil coronagraphs. *Astrophys. J.* **582**, 1147–1161 (2003).
8. Vanderbei, R. J., Spergel, D. N. & Kasdin, N. J. Circularly symmetric apodization via star-shaped masks. *Astrophys. J.* **599**, 686–694 (2003).
9. Cash, W., Kasdin, J., Seager, S. & Arenberg, J. Direct studies of exo-planets with the New Worlds Observer. *Proc. Soc. Photo-Opt. Instrum. Eng.* **5899**, 274–285 (2005).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** I thank R. Vanderbei, J. Kasdin, M. Lieber, and J. Arenberg for advice, discussion and encouragement. I thank D. Feldkhun, A. Lo, N. Rajan, E. Schindhelm, and W. Simmons for investigating alternative approaches and verifying the results of this paper. I wish to thank R. Cassanova and the NASA Institute for Advanced Concepts for financial support and encouragement of the work from its early stages.

**Author Information** Reprints and permissions information is available at [npg.nature.com/reprintsandpermissions](http://npg.nature.com/reprintsandpermissions). The author declares no competing financial interests. Correspondence and requests for materials should be addressed to W.C. ([cash@casa.colorado.edu](mailto:cash@casa.colorado.edu)).



## LETTERS

# Direct observation of the superfluid phase transition in ultracold Fermi gases

Martin W. Zwierlein<sup>1</sup>, Christian H. Schunck<sup>1</sup>, André Schirotzek<sup>1</sup> & Wolfgang Ketterle<sup>1</sup>

Phase transitions are dramatic phenomena: water freezes into ice, atomic spins spontaneously align in a magnet, and liquid helium becomes superfluid. Sometimes, such a drastic change in behaviour is accompanied by a visible change in appearance. The hallmark of Bose–Einstein condensation and superfluidity in trapped, weakly interacting Bose gases is the sudden formation of a dense central core inside a thermal cloud<sup>1–7</sup>. However, in strongly interacting gases—such as the recently observed fermionic superfluids<sup>8</sup>—there is no longer a clear separation between the superfluid and the normal parts of the cloud. The detection of fermion pair condensates has required magnetic field sweeps<sup>9–11</sup> into the weakly interacting regime, and the quantitative description of these sweeps presents a major theoretical challenge. Here we report the direct observation of the superfluid phase transition in a strongly interacting gas of <sup>6</sup>Li fermions, through sudden changes in the shape of the clouds—in complete analogy to the case of weakly interacting Bose gases. By preparing unequal mixtures of the two spin components involved in the pairing<sup>12,13</sup>, we greatly enhance the contrast between the superfluid core and the normal component. Furthermore, the distribution of non-interacting excess atoms serves as a direct and reliable thermometer. Even in the normal state, strong interactions significantly deform the density profile of the majority spin component. We show that it is these interactions that drive the normal-to-superfluid transition at the critical population imbalance of  $70 \pm 5$  per cent (ref. 12).

The dramatic signature of Bose–Einstein condensation in weakly interacting gases in atom traps derives from a natural hierarchy of energy scales: the critical temperature for condensation,  $T_C \propto n^{2/3}$  at particle density  $n$ , is much larger than the chemical potential (divided by the Boltzmann constant  $k_B$ ) of a pure condensate,  $\mu \propto na$ , which measures the interaction strength between particles ( $a$  is the scattering length). Hence, for weak (repulsive) interactions ( $a > 0$ ,  $na^3 \ll 1$ ), the condensate is clearly distinguished from the cloud of uncondensed particles through its smaller size and higher density. However, as the interactions are increased, for example by tuning  $a$  using a Feshbach resonance, this hierarchy of energy scales breaks down, as  $\mu$  can now become comparable to  $k_B T_C$ . In Fermi gases with weak attractive interaction ( $a < 0$ ,  $n|a|^3 \ll 1$ ), the chemical potential is given by the Fermi energy  $E_F$  and will even far exceed the superfluid transition temperature  $k_B T_C \propto E_F e^{-\pi/2k_F|a|}$  (where  $k_F \propto n^{1/3}$  is the Fermi wave vector). Both the normal and the condensed cloud will here be of the same size and shape, dependent only on  $E_F$  and the trapping potential.

The phase transition from the normal to the superfluid state, although dramatic in its consequences, is thus not revealed by a major change in the appearance of the gas. Indeed, in strongly interacting Fermi gases no deviation from a normal cloud's shape has so far been detected, either in the unitary regime, where  $a$  diverges, or on the attractive Bardeen–Cooper–Schrieffer (BCS) side of a Feshbach resonance. Theoretical works predicted small

'kinks'<sup>14–16</sup> or other slight deviations<sup>17</sup> in the density profiles of the gas in the superfluid regime, but after line-of-sight integration these effects have so far been too small to be observable. Condensates could only be observed via rapid magnetic field ramps to the Bose–Einstein condensate (BEC) side ( $a > 0$ ) of the Feshbach resonance, performed during expansion<sup>9,10</sup>. This suddenly reduced the condensate's chemical potential, and let the thermal fraction grow beyond the condensate size. A similar ramp was used to detect vortices on resonance and on the BCS side in the demonstration of fermionic superfluidity<sup>8</sup>. However, these magnetic field ramps are difficult to model theoretically, and a satisfactory quantitative comparison of, for example, the condensate fraction with experiments has not been accomplished<sup>18–21</sup>.

In this work we demonstrate that the normal-to-superfluid phase transition in a strongly interacting Fermi gas can be directly observed in absorption profiles, without the need for any magnetic field ramps. As in the case of weakly interacting BECs, preparation, expansion and detection of the sample all take place at the same, fixed magnetic field and scattering length. As for BECs, the phase transition is observed as a sudden change in the shape of the cloud during time-of-flight expansion, when the trap depth is decreased below a critical value. To clearly distinguish the superfluid from the normal component, we break the number symmetry between spin-up (majority atom number,  $N_\uparrow$ ) and spin-down (minority atom number,  $N_\downarrow$ ) and produce an unequal mixture of fermions (imbalance parameter  $\delta = (N_\uparrow - N_\downarrow)/(N_\uparrow + N_\downarrow)$ ). Standard BCS superfluidity requires equal densities of the two spin components. Hence, when cooled below the phase transition the cloud should show a sudden onset of a superfluid region of equal densities. Indeed, below a critical temperature, we observe how the density distribution of the minority component becomes bimodal.

Breaking the symmetry in atom numbers thus produces a direct and striking signature of the superfluid phase transition<sup>22–24</sup>. A similar situation has been encountered in Bose–Einstein condensation, where breaking the symmetry of a spherical trap resulted in dramatic anisotropic expansion of the condensate, now a hallmark of the BEC phase transition.

Figure 1 shows column density profiles of the two imbalanced spin states for different points along the evaporation path corresponding to different temperatures, and for three magnetic fields that correspond to the BEC side, exact resonance and the BCS side of the resonance. For large final trap depths (upper panels in Fig. 1), the smaller cloud has the expected shape of a normal, non-superfluid gas: it is very well fitted using a single, finite temperature Thomas–Fermi-profile (with central optical density, radius and the fugacity as independent fit-parameters). However, below a critical trap depth, a second, denser feature appears in the centre of the minority component (lower panels in Fig. 1). This onset of bimodality occurs very suddenly as the trap depth is lowered, as can be seen from Fig. 2: Around the critical point, the atom number (Fig. 2a) and population

<sup>1</sup>Department of Physics, MIT-Harvard Center for Ultracold Atoms, and Research Laboratory of Electronics, MIT, Cambridge, Massachusetts 02139, USA.

imbalance (Fig. 2b) are practically constant, and the temperature (Fig. 2c) varies in a smooth linear way with the trap depth. In contrast, below the critical trap depth, the shape of the smaller cloud starts to deviate drastically from the Thomas–Fermi distribution of a normal gas, as quantified in Fig. 2d. This sudden increase in the standard deviation of a fit to a single-component fitting function is a standard way of identifying the BEC phase transition in a model-independent way<sup>2</sup>.

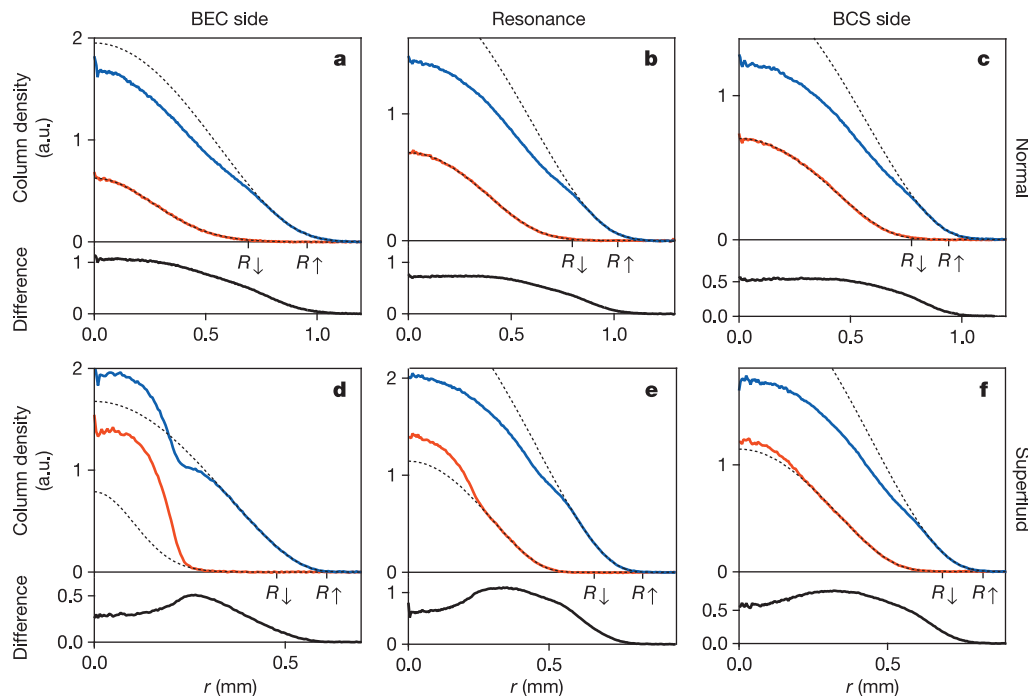
Figure 2e displays the fact that below the critical trap depth a new, third radius is required to describe the two clouds. As we will see below, the appearance of this central feature coincides with the appearance of the fermion pair condensate in experiments involving the magnetic field ramp technique<sup>10–12</sup>. It is this condensate that contains the superfluid vortices in refs 8 and 12. We are thus naturally led to interpret the central core as the condensate of fermion pairs, and the outer wings as the normal, uncondensed part of the cloud. This constitutes, to our knowledge, the first direct observation of the normal-to-superfluid phase transition in resonantly interacting Fermi gases on resonance and on the BCS side (that is, without a magnetic field sweep that so far cannot be quantitatively accounted for).

Already at high temperatures, above the phase transition, the larger cloud's profile is strongly deformed in the presence of the smaller cloud, a direct signature of interaction. Indeed, on resonance the cloud size of the minority component is significantly smaller than that of a non-interacting sample with the same number of atoms (see Fig. 2e). At the phase transition, the outer radii of the clouds do not change abruptly. This demonstrates that interactions, not superfluidity, are the main mechanism behind the reduced cloud size of an interacting Fermi gas.

On the BEC side, the condensate is clearly visible in the larger cloud.

On resonance, however, the condensate is not easily discernible in the larger component's profiles at the scale of Fig. 1. Nevertheless, we have found a very faint but reproducible trace of the condensate when analysing the curvature of these column density profiles (see Supplementary Fig. S1). On resonance and on the BCS side, the onset of bimodality in the smaller cloud can be clearly observed for imbalances larger than  $\sim 20\%$  (but below a certain critical imbalance, see below), for which the condensate is small compared to the minority cloud size. With increasing magnetic field on the BCS side (that is, with decreasing interaction strength), the bimodality becomes less pronounced and is not clearly discerned beyond 853 G (interaction parameter  $1/k_F a < -0.15$ ).

Thermometry of strongly interacting Fermi gases has always been a major difficulty in experiments on strongly interacting fermions<sup>25</sup>. A thermometer can only be reliable if the working substance is not affected by the sample to be measured. In equal mixtures of fermions, the two overlapping atomic clouds are strongly interacting throughout. Temperatures determined from a non-interacting Thomas–Fermi fit to these clouds need calibration based on approximate theoretical calculations<sup>25</sup>. In addition, as will be reported elsewhere, we find that those fits do not describe the profiles of a partially superfluid Fermi gas as well as they do in the normal state, in agreement with theory<sup>14–17</sup>. In the case of imbalanced mixtures, the wings of the larger component, where the spin-down species are absent, are non-interacting and thus serve as a direct thermometer (see Fig. 2c). For an imbalance of  $\delta = 75 \pm 3\%$  we determine the critical temperature for the phase transition on the BEC side at  $1/k_F a = 0.46$  to be  $T/T_F = 0.18(3)$  ( $k_B T_F = \hbar \omega(3(N_\uparrow + N_\downarrow))^{1/3} \equiv \hbar^2 k_F^2/2m$  is the Fermi energy of a non-interacting, equal mixture with the same total number of fermions  $N_\uparrow + N_\downarrow$ ,  $\omega/2\pi$  is the



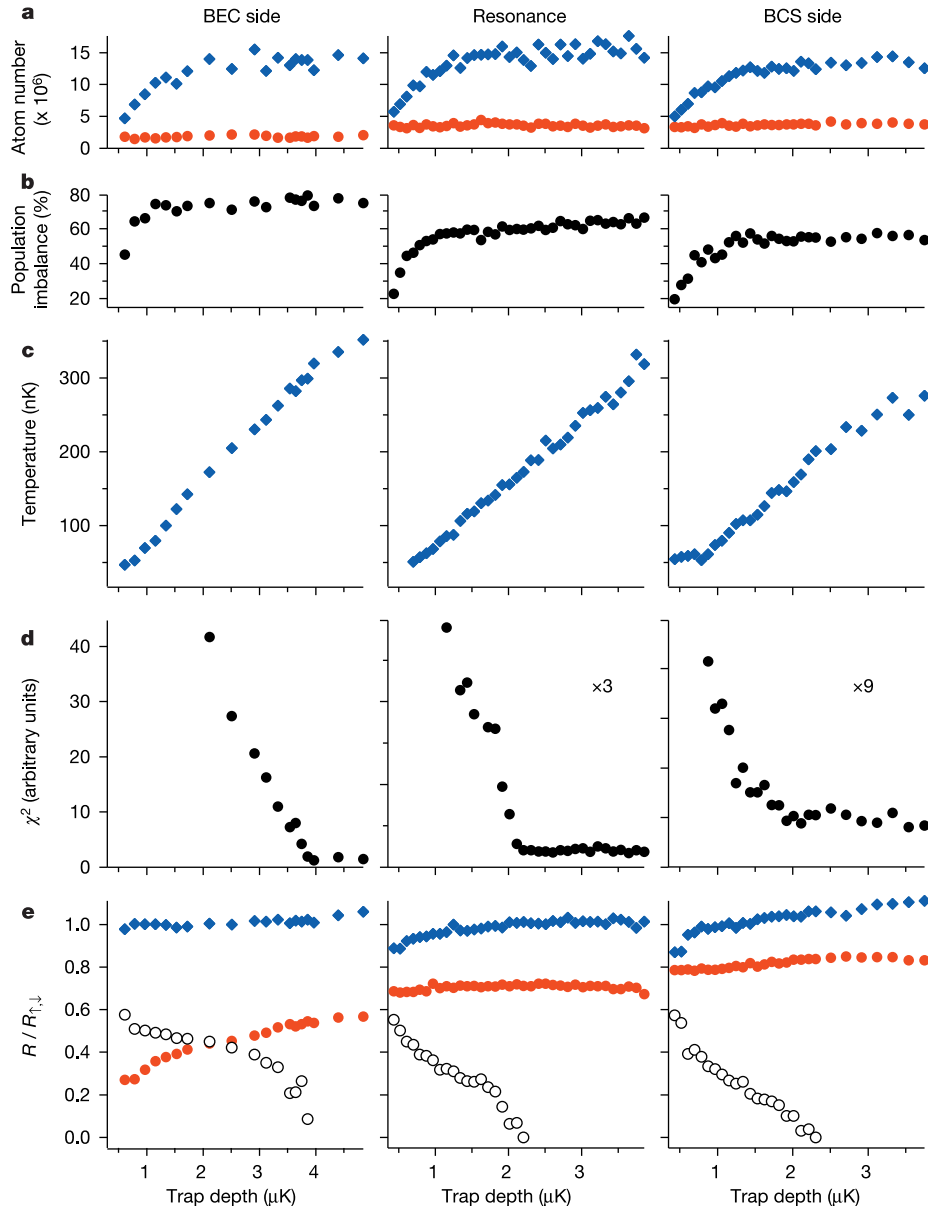
**Figure 1 | Direct observation of the phase transition in a strongly interacting two-state mixture of fermions with imbalanced spin populations.** Top (a–c) and bottom (d–f) rows show the normal and the superfluid state, respectively. Panels a and d were obtained in the BEC regime (at  $B = 781$  G), b and e on resonance (834 G), and c and f on the BCS side of the Feshbach resonance (853 G). The profiles represent the azimuthal average of the column density after 10 ms (BEC side) or 11 ms (on resonance and BCS side) of expansion. The appearance of a dense central feature in the smaller component marks the onset of condensation. The condensate causes a clear depletion in the difference profiles (bottom of each panel). Both in the normal and in the superfluid state, interactions between the two spin

states are manifest in the strong deformation of the larger component. The dotted lines show Thomas–Fermi fits to the wings of the column density. The radii  $R_\uparrow$  and  $R_\downarrow$  mark the Fermi radius of a ballistically expanding, non-interacting cloud with atom number  $N_\uparrow$ ,  $N_\downarrow$ . The trap depth  $U$  (in  $\mu\text{K}$ ), the atom numbers, the population imbalance  $\delta$  (in %), the interaction parameter  $1/k_F a$ , the temperature  $T$  (in nK) and the reduced temperature  $T/T_F$  were respectively: a, 4.8,  $1.8 \times 10^7$  and  $2.6 \times 10^6$ , 75, 0.42, 350, 0.20; b, 3.2,  $1.8 \times 10^7$  and  $4.2 \times 10^6$ , 63, 0 (resonance), 260, 0.15; c, 2.5,  $1.5 \times 10^7$  and  $4.5 \times 10^6$ , 52,  $-0.13$ , 190, 0.12; d, 0.8,  $6.5 \times 10^6$  and  $1.5 \times 10^6$ , 62, 0.67, 50,  $\leq 0.05$ ; e, 1.1,  $1.5 \times 10^7$  and  $3.8 \times 10^6$ , 60%, 0 (resonance), 70, 0.06; f, 1.2,  $1.3 \times 10^7$  and  $4.4 \times 10^6$ , 50,  $-0.15$ , 100, 0.08. a.u., arbitrary units.

geometric mean of the trapping frequencies, and  $m$  is the mass of  $^6\text{Li}$ . This corresponds to  $T/T_{\text{C},\downarrow} = 0.55(9)$  when comparing the temperature to the critical temperature  $T_{\text{C},\downarrow}$  for Bose condensation in a non-interacting gas with  $N_{\downarrow}$  bosons. The reduction in the critical temperature is a direct consequence of strong repulsive interactions between the molecules. On resonance, at  $\delta = 59 \pm 3\%$ , we find  $T/T_{\text{F}} = 0.12(2)$ , and on the BCS side ( $1/k_{\text{F}}a = -0.14$ ) for  $\delta = 53 \pm 3\%$  we obtain  $T/T_{\text{F}} = 0.11(2)$ . These are, to our knowledge, the first directly measured and reliable temperatures for the superfluid transition in strongly interacting Fermi gases. They may serve as a checkpoint for theoretical models.

We note that the critical temperature will in general depend on the population imbalance. For example, for large enough imbalance on resonance or on the BCS side, no condensate will form even at zero temperature<sup>12</sup>, as we discuss below. Here, the critical temperature for superfluidity will be zero.

An important qualitative difference distinguishes the BEC side from resonance at the lowest temperatures. On the BEC side, the gas consists of only two parts—the superfluid core surrounded by a fully polarized degenerate Fermi gas of the excess species. On resonance and on the BCS side, however, there exists a third region, a normal state in which both species are mixed. Several recent theories describe



**Figure 2 | Characterization of the phase transition.** **a–e**, The data characterize the evolution of the fermion mixture as the cloud is evaporatively cooled by lowering the trap depth. The chosen magnetic fields are identical to those in Fig. 1. Data obtained from the majority (minority) cloud are shown as diamonds (circles). **a**, The atom number; **b**, the population imbalance between the two spin states; and **c**, the temperature of the spin mixture as determined from the non-interacting wings of the larger cloud's profile. **d**, A finite temperature Fermi–Dirac (for resonance and the BCS side) or gaussian (for the BEC side) distribution is fitted to the minority cloud; the phase transition is marked by a sudden increase in  $\chi^2$  as the condensate starts to appear. **e**, Outer radii of the majority and

minority cloud (for the minority cloud on the BEC side, thermal cloud radius; all other cases, Thomas–Fermi radius) as well as the condensate radius (open circles), defined as the position of the ‘kink’ in the minority profile (see Fig. 1). The majority (minority and condensate) cloud size is normalized by the Fermi radius  $R_{\uparrow}$  ( $R_{\downarrow}$ ) of a non-interacting cloud with  $N_{\uparrow}$  ( $N_{\downarrow}$ ) atoms, and adjusted for ballistic (hydrodynamic) expansion. Note that the imbalance decreases during evaporation because the larger majority cloud incurs stronger evaporative losses. For the data, three (BEC and resonance) to five (BCS) independent measurements were averaged.



density profiles of imbalanced Fermi mixtures<sup>26–32</sup>. Mean-field theories that neglect interactions in the normal cloud and between the normal and condensed cloud are only in qualitative agreement with our results. Descriptions that exclude the mixed region or find superfluidity on resonance at all population imbalances are ruled out by our observations.

To elucidate the origin of the clear separation between condensate and normal components, we varied the population imbalance at our coldest temperatures and on resonance. Figure 3b shows several resulting profiles after 11 ms expansion from the trap. For large imbalances,  $\delta > 70\%$ , the minority cloud is not bimodal and well fitted by a (unconstrained) Thomas–Fermi profile. At a critical imbalance of  $\delta \approx 70\%$ , the condensate appears and then grows further as the imbalance is reduced (for the cloud radii, see Supplementary Fig. S2).

To characterize the appearance of the condensate for imbalances around  $\delta = 70\%$ , a Thomas–Fermi profile is fitted to the wings of the minority cloud. The fraction of atoms not contained in this fit is a measure of the condensate fraction (see Fig. 3). We find a critical imbalance of  $\delta_c = 70(5)\%$  above which the condensate disappears. This agrees with our previous work<sup>12</sup>, where we employed a rapid ramp method to the BEC side to extract the condensate fraction. We

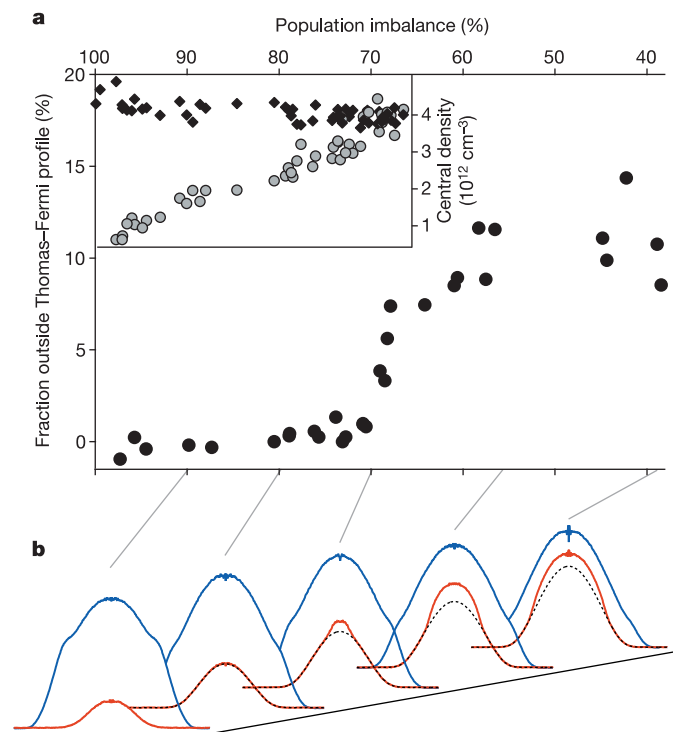
observed the quantum phase transition from the superfluid to the normal state as a critical population imbalance of  $\delta_c = 70\%$  was exceeded. This strongly suggests that the bimodality observed here directly in the minority component, and the bimodality observed in molecular clouds after a magnetic field sweep, are signatures of the same phase transition.

The transition at  $\delta_c$  is known as the Clogston limit of superfluidity<sup>12,33</sup>, and occurs when the chemical potential difference  $\delta\mu$  becomes larger than a constant times the (local) superfluid gap  $\Delta(\mathbf{r})$  (see Supplementary Information). Here we present a simple picture for the character of this phase transition in a harmonic trap. Thomas–Fermi fits for the normal clouds beyond  $\delta_c$  allow a simple estimate of the central three-dimensional density of the gas (with an estimated accuracy of 20% for the relative density difference), shown in the inset of Fig. 3. For large imbalances, we find that the three-dimensional densities differ significantly, as is expected for two weakly interacting Fermi clouds. As the imbalance is reduced towards the critical  $\delta_c$ , the central densities approach each other and become approximately equal around  $\delta_c$ . This is a direct consequence of strong interactions in the normal state. In a non-interacting Fermi mixture with an imbalance of  $\delta_c$ , the central densities would differ by a factor of 2.4.

This observation now offers an intriguing insight into the nature of a fermionic superfluid on resonance or on the BCS side. Already in the normal state above  $T_c$  or beyond  $\delta = \delta_c$ , interactions between the two spin states are strong. Indeed, this is directly seen in the deformation of the majority cloud due to the presence of the minority species (see Figs 1, 3). However, here these interactions are not strong enough to let the central densities of the two clouds become comparable. At the critical imbalance the Clogston criterion  $\delta\mu = c\Delta(\mathbf{r} = 0)$  is fulfilled in the centre of the trap (here,  $c$  is a constant that equals  $\sqrt{2}$  in the BCS limit<sup>33</sup>). For smaller imbalance, a central superfluid region can form: the condensate. Its borders are defined by  $\delta\mu < c\Delta(\mathbf{r})$ . The simple density estimate in Fig. 3 suggests that in this region, the two clouds will have equal densities, although more refined techniques to measure small density differences have to be developed to finally settle this question. Outside the superfluid region there is still a normal state with unequal densities of minority and majority components. The discontinuity in the clouds' densities at the normal-to-superfluid phase boundary gives rise to the visible kink in the column density profiles. Such a density discontinuity is characteristic of a first-order phase transition.

Interestingly, most of the 'work' needed to build the superfluid state has already been done in the normal component by decreasing the density difference. Consequently, the critical population difference needed to form the superfluid is largely determined by the interactions in the normal gas.

In conclusion, we have observed the normal-to-superfluid phase transition through the direct observation of condensation in an imbalanced Fermi mixture—on the BEC side, on the BCS side, and right on the Feshbach resonance. Unequal mixtures offer a direct method of thermometry by analysing the non-interacting wings of the majority species. Strong interactions are already visible in the normal cloud as marked deformations of the majority profile. It is these interactions in the normal gas that squeeze the two components and eventually, at the critical imbalance, let them reach almost equal densities in the centre, aiding the formation of the superfluid. Our method of direct detection of the condensate is a powerful new tool to characterize the superfluid phase transition. At the current level of precision, the appearance of a condensate after magnetic field sweeps and the direct observation of the central dense core occur together, and indicate the normal-to-superfluid phase transition. An intriguing question is whether further phases are possible, including a more exotic superfluid state with unequal densities. Several theories predict that the Fulde–Ferrell–Larkin–Ovchinnikov state, a superfluid state with oscillating order parameter, should be present for imbalanced spin populations<sup>24,26,28</sup>.



**Figure 3 | Quantum phase transition to superfluidity for decreasing population imbalance.** **a**, Main panel, the ‘condensate fraction’ of excess minority atoms, not contained in the Thomas–Fermi fit, versus population imbalance on resonance. **b**, Column density profiles of majority (blue) and minority (red) clouds, azimuthally averaged, for varying population imbalance. The condensate is clearly visible in the minority component as the dense central feature on top of the normal background (finite-temperature Thomas–Fermi fit, dotted lines). Below the critical imbalance  $\delta_c = 70\%$ , the condensate starts to form. The inset in **a** shows the central densities of the larger (black diamonds) and smaller (grey circles) cloud in the normal state above  $\delta_c$ . This demonstrates that here the central densities are unequal, suppressing superfluidity. The densities were calculated from the central optical density and the fitted size of the clouds, assuming local density approximation and adjusting for ballistic (hydrodynamic) expansion of the outer radii of majority (minority) clouds. The data points for the condensate fraction show the average of several independent measurements.

## METHODS

**Experimental procedure.** Our experimental setup is described in previous publications<sup>8,12</sup>. A spin-polarized cloud of <sup>6</sup>Li fermions is cooled to degeneracy using a combination of laser cooling and sympathetic cooling with sodium atoms in a magnetic trap. After transfer into an optical trap, a variable spin mixture of the lowest two hyperfine states, labelled  $|\uparrow\rangle$  and  $|\downarrow\rangle$ , is prepared at a magnetic bias field of 875 G. Interactions between the two spin states can be freely tuned via a 300-G-wide Feshbach resonance located at  $B_0 = 834$  G. At fields below  $B_0$ , two-body physics supports a stable molecular bound state (BEC side), while at higher fields (BCS side), no such bound state exists for two isolated atoms. Our trap combines a magnetic saddle potential with a weakly focused (waist  $w \approx 120 \mu\text{m}$ ) infrared laser beam (wavelength  $\lambda = 1,064 \text{ nm}$ ), leading to a harmonic axial confinement with oscillation frequency of  $\nu_z = 22.8(0.2) \text{ Hz}$  and a gaussian radial potential with variable trapping frequency  $\nu_r$  in the central harmonic region. The trap depth  $U$  is related to  $\nu_r$  and  $\nu_z$  by:

$$U = \frac{1}{4} m (2\pi\nu_r)^2 w^2 \left( 1 - \frac{\nu_z^2}{2\nu_r^2} \ln \left( \frac{2(\nu_r^2 + \nu_z^2/2)}{\nu_z^2} \right) \right).$$

The initial degeneracy of the spin mixture is about  $T/T_F \approx 0.3$ . The strongly interacting gas is further cooled by decreasing the laser power of the optical trap over several seconds and evaporating the most energetic particles. During the first few seconds, the magnetic field is adiabatically ramped to a chosen final field in the resonance region where the last stage of the evaporation (shown in Fig. 2) takes place. For detection, the optical trap is switched off and the gas expands in the remaining magnetic saddle-point potential. After a variable time-of-flight, an absorption image of atoms either in state  $|\uparrow\rangle$  or  $|\downarrow\rangle$  is taken along the axial direction of the trap (the direction of the optical trapping beam). The cloud's radial symmetry allows for azimuthal averaging of the resulting column densities, leading to low-noise profiles<sup>12</sup>.

For preparing clouds at the coldest temperatures (as shown in Fig. 3) with varying population imbalance, the spin mixture is evaporated down to a trap depth of  $1 \mu\text{K}$  over several seconds on resonance, after which the trap depth is increased again to  $1.4 \mu\text{K}$  for more harmonic confinement (trap frequencies:  $\nu_r = 115(10) \text{ Hz}$  and  $\nu_z = 22.8(0.2) \text{ Hz}$ ). The temperature of the gas is determined to be  $T/T_F \leq 0.06$  for all  $\delta > 15\%$ , and appears to smoothly rise to  $T/T_F = 0.11$  for an equal mixture, although thermometry in the interacting wings is problematic. The total atom number was  $1.5 \times 10^7$  and constant to within 15% for all values of  $\delta$ .

**Errors.** The error in the critical temperature  $T_C/T_F$  for the phase transition is dominated by the uncertainty in the atom number entering the determination of  $T_F$ , which we estimate to be 30% (ref. 12). For  $T_F$  we use the harmonic approximation for the radially gaussian trapping potential, with the measured trapping frequencies reflecting the average curvature of the gaussian potential. The phase transition is observed above  $U = 2 \mu\text{K}$ , where anharmonicities contribute only 3% to the error in  $T_F$ . Note that anharmonicities do not affect the temperature measurement performed on the majority wings: ballistic expansion of non-interacting atoms reveals their momentum distribution, regardless of the shape of the trap.

Received 6 March; accepted 23 May 2006.

- Anderson, M. H., Ensher, J. R., Matthews, M. R., Wieman, C. E. & Cornell, E. A. Observation of Bose-Einstein condensation in a dilute atomic vapor. *Science* **269**, 198–201 (1995).
- Davis, K. B. *et al.* Bose-Einstein condensation in a gas of sodium atoms. *Phys. Rev. Lett.* **75**, 3969–3973 (1995).
- Greiner, M., Regal, C. A. & Jin, D. S. Emergence of a molecular Bose-Einstein condensate from a Fermi gas. *Nature* **426**, 537–540 (2003).
- Zwierlein, M. W. *et al.* Observation of Bose-Einstein condensation of molecules. *Phys. Rev. Lett.* **91**, 250401 (2003).
- Bartenstein, M. *et al.* Crossover from a molecular Bose-Einstein condensate to a degenerate Fermi gas. *Phys. Rev. Lett.* **92**, 120401 (2004).
- Bourdel, T. *et al.* Experimental study of the BEC-BCS crossover region in lithium 6. *Phys. Rev. Lett.* **93**, 050401 (2004).
- Partridge, G. B., Strecker, K. E., Kamar, R. I., Jack, M. W. & Hulet, R. G. Molecular probe of pairing in the BEC-BCS crossover. *Phys. Rev. Lett.* **95**, 020404 (2005).

- Zwierlein, M. W., Abo-Shaeer, J. R., Schirotzek, A., Schunck, C. H. & Ketterle, W. Vortices and superfluidity in a strongly interacting Fermi gas. *Nature* **435**, 1047–1051 (2005).
- Regal, C. A., Greiner, M. & Jin, D. S. Observation of resonance condensation of fermionic atom pairs. *Phys. Rev. Lett.* **92**, 040403 (2004).
- Zwierlein, M. W. *et al.* Condensation of pairs of fermionic atoms near a Feshbach resonance. *Phys. Rev. Lett.* **92**, 120403 (2004).
- Zwierlein, M. W., Schunck, C. H., Stan, C. A., Raupach, S. M. F. & Ketterle, W. Formation dynamics of a fermion pair condensate. *Phys. Rev. Lett.* **94**, 180401 (2005).
- Zwierlein, M. W., Schirotzek, A., Schunck, C. H. & Ketterle, W. Fermionic superfluidity with imbalanced spin populations. *Science* **311**, 492–496 (2006). Published online 21 December 2005 (doi:10.1126/science.1122318).
- Partridge, G. B., Li, W., Kamar, R. I., Liao, Y. & Hulet, R. G. Pairing and phase separation in a polarized Fermi gas. *Science* **311**, 503–505 (2006). Published online 21 December 2005 (doi:10.1126/science.1122876).
- Chiofalo, M. L., Kokkelmans, S. J. J. M. F., Milstein, J. N. & Holland, M. J. Signatures of resonance superfluidity in a quantum Fermi gas. *Phys. Rev. Lett.* **88**, 090402 (2002).
- Ho, T.-L. Universal thermodynamics of degenerate quantum gases in the unitarity limit. *Phys. Rev. Lett.* **92**, 090402 (2004).
- Perali, A., Pieri, P., Pisani, L. & Strinati, G. C. BCS-BEC crossover at finite temperature for superfluid trapped Fermi atoms. *Phys. Rev. Lett.* **92**, 220404 (2004).
- Stajic, J., Chen, Q. & Levin, K. Density profiles of strongly interacting trapped Fermi gases. *Phys. Rev. Lett.* **94**, 060401 (2005).
- Diener, R. B. & Ho, T.-L. Projecting fermion pair condensates into molecular condensates. Preprint at (<http://arxiv.org/cond-mat/0404517>) (2004).
- Perali, A., Pieri, P. & Strinati, G. C. Extracting the condensate density from projection experiments with Fermi gases. *Phys. Rev. Lett.* **95**, 010407 (2005).
- Altman, E. & Vishwanath, A. Dynamic projection on Feshbach molecules: A probe of pairing and phase fluctuations. *Phys. Rev. Lett.* **95**, 110404 (2005).
- Chen, Q., Regal, C. A., Greiner, M., Jin, D. S. & Levin, K. Understanding the superfluid phase diagram in trapped Fermi gases. *Phys. Rev. A* **73**, 041603 (2006).
- Bedaque, P. F., Caldas, H. & Rupak, G. Phase separation in asymmetrical fermion superfluids. *Phys. Rev. Lett.* **91**, 247002 (2003).
- Caldas, H. Cold asymmetrical fermion superfluids. *Phys. Rev. A* **69**, 063602 (2004).
- Sheehy, D. E. & Radzihovsky, L. BEC-BCS crossover in “magnetized” Feshbach-resonantly paired superfluids. *Phys. Rev. Lett.* **96**, 060401 (2006).
- Kinast, J. *et al.* Heat capacity of a strongly-interacting Fermi gas. *Science* **307**, 1296–1299 (2005).
- Mizushima, T., Machida, K. & Ichioka, M. Direct imaging of spatially modulated superfluid phases in atomic fermion systems. *Phys. Rev. Lett.* **94**, 060404 (2005).
- Pieri, P. & Strinati, G. C. Trapped fermions with density imbalance in the Bose-Einstein condensate limit. *Phys. Rev. Lett.* **96**, 150404 (2006).
- Kinnunen, J., Jensen, L. M. & Törmä, P. Strongly interacting Fermi gases with density imbalance. *Phys. Rev. Lett.* **96**, 110403 (2006).
- De Silva, T. N. & Mueller, E. J. Profiles of near-resonant population-imbalanced trapped Fermi gases. *Phys. Rev. A* **73**, 051602(R) (2006).
- Yi, W. & Duan, L.-M. Trapped fermions across a Feshbach resonance with population imbalance. *Phys. Rev. A* **73**, 031604(R) (2006).
- Chevy, F. Density profile of a trapped strongly interacting Fermi gas with unbalanced spin populations. *Phys. Rev. Lett.* **96**, 130401 (2006).
- Haque, M. & Stoof, H. T. C. Pairing of a trapped resonantly-interacting fermion mixture with unequal spin populations. Preprint at (<http://arxiv.org/cond-mat/0601321>) (2006).
- Clogston, A. M. Upper limit for the critical field in hard superconductors. *Phys. Rev. Lett.* **9**, 266–267 (1962).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank the participants of the Aspen winter conference on strongly interacting fermions for discussions. This work was supported by the NSF, ONR and NASA.

**Author Information** Reprints and permissions information is available at [npg.nature.com/reprintsandpermissions](http://npg.nature.com/reprintsandpermissions). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.W.Z. (zwierlein@mit.edu).

# Resonance in the electron-doped high-transition-temperature superconductor $\text{Pr}_{0.88}\text{LaCe}_{0.12}\text{CuO}_{4-\delta}$

Stephen D. Wilson<sup>1</sup>, Pengcheng Dai<sup>1,2</sup>, Shiliang Li<sup>1</sup>, Songxue Chi<sup>1</sup>, H. J. Kang<sup>3,4</sup> & J. W. Lynn<sup>3</sup>

In conventional superconductors, the interaction that pairs the electrons to form the superconducting state is mediated by lattice vibrations<sup>1</sup> (phonons). In high-transition-temperature (high- $T_c$ ) copper oxides, it is generally believed that magnetic excitations might play a fundamental role in the superconducting mechanism because superconductivity occurs when mobile 'electrons' or 'holes' are doped into the antiferromagnetic parent compounds<sup>2</sup>. Indeed, a sharp magnetic excitation termed 'resonance' has been observed by neutron scattering in a number of hole-doped materials<sup>3–11</sup>. The resonance is intimately related to superconductivity<sup>12</sup>, and its interaction with charged quasi-particles observed by photoemission<sup>13,14</sup>, optical conductivity<sup>15</sup>, and tunnelling<sup>16</sup> suggests that it might play a part similar to that of phonons in conventional superconductors. The relevance of the resonance to high- $T_c$  superconductivity, however, has been in doubt because so far it has been found only in hole-doped materials<sup>17</sup>. Here we report the discovery of the resonance in electron-doped superconducting  $\text{Pr}_{0.88}\text{LaCe}_{0.12}\text{CuO}_{4-\delta}$  ( $T_c = 24$  K). We find that the resonance energy ( $E_r$ ) is proportional to  $T_c$  via  $E_r \approx 5.8k_B T_c$  for all high- $T_c$  superconductors irrespective of electron- or hole-doping. Our results demonstrate that the resonance is a fundamental property of the superconducting copper oxides and therefore must be essential in the mechanism of superconductivity.

Although the interaction of electrons with phonons or magnetic excitations can cause electron pairing and superconductivity, we focus on magnetic excitations because the resonance is intimately related to superconductivity and is also present in several classes of hole-doped high- $T_c$  materials. The resonance is a sharp magnetic excitation centred at the wavevector  $\mathbf{Q} = (1/2, 1/2)$  in the two-dimensional reciprocal space of the  $\text{CuO}_2$  planes, which corresponds to the antiferromagnetic Bragg position of the undoped compounds (Fig. 1a). It was first discovered in the hole-doped bilayer (each lattice unit cell has two  $\text{CuO}_2$  planes) high- $T_c$  superconductor  $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$  (YBCO)<sup>3</sup>. Its intensity grows below  $T_c$  and its energy ( $\hbar\omega$ ) scales approximately with  $k_B T_c$  (Fig. 1e)<sup>4–8</sup>. At energies below the resonance, spin fluctuations peak at incommensurate wavevectors<sup>8</sup> and disperse inwards towards the resonance<sup>18,19</sup>. Such behaviour is remarkably similar to that of the hole-doped  $\text{La}_{2-x}(\text{Ba},\text{Sr})_x\text{CuO}_4$  (refs 20–22). Although the resonance has also been observed in the hole-doped bilayer  $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$  (called Bi(2212))<sup>9,10</sup> and in the single-layer  $\text{Tl}_2\text{Ba}_2\text{CuO}_{6+\delta}$ <sup>11</sup>, the wavevector and energy dependence of the mode has been determined only for YBCO<sup>3–8</sup> because of the small available single crystal volumes in the Bi(2212)<sup>9,10</sup> and  $\text{Tl}_2\text{Ba}_2\text{CuO}_{6+\delta}$ <sup>11</sup>. Therefore, it has not been possible directly and systematically to compare the neutron results with those of the photoemission, which are obtained mostly for Bi(2212)<sup>13,14</sup>.

Here we studied the magnetic excitations of electron-doped materials to understand the electron-hole symmetry in high- $T_c$

superconductors. We grew large single crystals of  $\text{Pr}_{0.88}\text{LaCe}_{0.12}\text{CuO}_{4-\delta}$  (PLCCO) using the travelling solvent floating zone technique and annealed the samples to obtain optimal superconductivity with  $T_c = 24$  K (Fig. 1a)<sup>23,24</sup>. PLCCO is a single-layer electron-doped copper oxide and was chosen to avoid the static antiferromagnetic order that coexists with superconductivity in  $\text{Nd}_{1.85}\text{Ce}_{0.15}\text{CuO}_4$  (NCCO)<sup>25</sup>; the  $T_c = 24$  K PLCCO is phase pure, without static antiferromagnetic order<sup>23,26</sup>. Our elastic  $\mathbf{Q}$ -scans through the expected antiferromagnetic Bragg positions confirmed no static antiferromagnetic order to at least 600 mK in our PLCCO samples (Figs 1b and 1c).

We first probed the low-energy magnetic excitations of PLCCO using the Spin-Polarized Inelastic Neutron-Scattering Spectrometer (SPINS). The magnetic excitations were commensurate and centred around  $\mathbf{Q} = (1/2, 1/2, 0)$  at all temperatures (Figs 2a–c), similar to those of NCCO<sup>25</sup>. They can be well described by gaussians on linear backgrounds and are not resolution-limited. Fourier transforms of the gaussian peaks in Figs 2a–c gave dynamic spin correlation lengths of  $\xi \approx 96 \pm 15$ ,  $80 \pm 10$  and  $94 \pm 24$  Å for  $\hbar\omega = 0.5$ , 1.5 and 3.5 meV, respectively. With increasing temperature, the scattering above background was virtually unchanged from 2 K ( $T_c - 22$  K) to 30 K ( $T_c + 6$  K) but decreased slightly at 55 K owing to increased background (Figs 2a–c). The temperature dependence of the scattering at  $\hbar\omega = 1.5$  meV showed no observable anomaly across  $T_c$  (Fig. 2d), thus suggesting that the magnetic excitations are gapless and different from those of NCCO<sup>25</sup>. The energy scans at  $\mathbf{Q} = (1/2, 1/2, 0)$  confirmed that the magnetic scattering between 0.5 and 4.5 meV is virtually temperature independent between 2 and 30 K (Fig. 2e), and therefore does not follow the population factor  $1/(1 - e^{-\hbar\omega/k_B T})$  expected for simple bosonic excitations.

To study the magnetic excitations for energies above 4.5 meV, we used the HB-1 and BT-9 thermal neutron triple-axis spectrometers. Figure 3a–c shows  $\mathbf{Q}$ -scans through  $(1/2, 1/2, 0)$  for  $\hbar\omega = 3.5$ , 8.0 and 10 meV at  $T = 2$ , 30 and 80 K. In contrast to the temperature-independent low-energy ( $\hbar\omega \leq 4.5$  meV) magnetic excitations below  $T_c$  (Fig. 2), the integrated intensity above background around  $(1/2, 1/2, 0)$  at  $\hbar\omega = 8$  and 10 meV shows a significant enhancement on cooling from 30 K to 2 K, but hardly changes on warming from 30 K to 80 K (Fig. 3b and c). The  $\mathbf{Q}$ -width of the peak at  $\hbar\omega = 10$  meV is temperature-independent and resolution-limited, giving a minimum  $\xi \approx 45 \pm 5$  Å (Fig. 3c). Similar scans using better collimations on BT-9 (Fig. 4e) are again resolution-limited, giving  $\xi \approx 51 \pm 7$  Å. Figure 3d shows energy scans at the peak centre [ $\mathbf{Q} = (1/2, 1/2, 0)$ ] compared to background [ $\mathbf{Q} = (0.5875, 0.5875, 0)$ ] positions for temperatures 2, 30 and 80 K. The weak, dispersion-less peak at  $\hbar\omega = 6$  meV and the gradual rising background scattering with increasing energy are due to the  $\text{Pr}^{3+}$  crystalline electric-field excitations in the tetragonal unit cell of PLCCO<sup>27</sup>. Its intensity simply

<sup>1</sup>Department of Physics and Astronomy, The University of Tennessee, Knoxville, Tennessee 37996-1200, USA. <sup>2</sup>Center for Neutron Scattering, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA. <sup>3</sup>NIST Center for Neutron Research, National Institute of Standards and Technology, Gaithersburg, Maryland 20899-8562, USA.

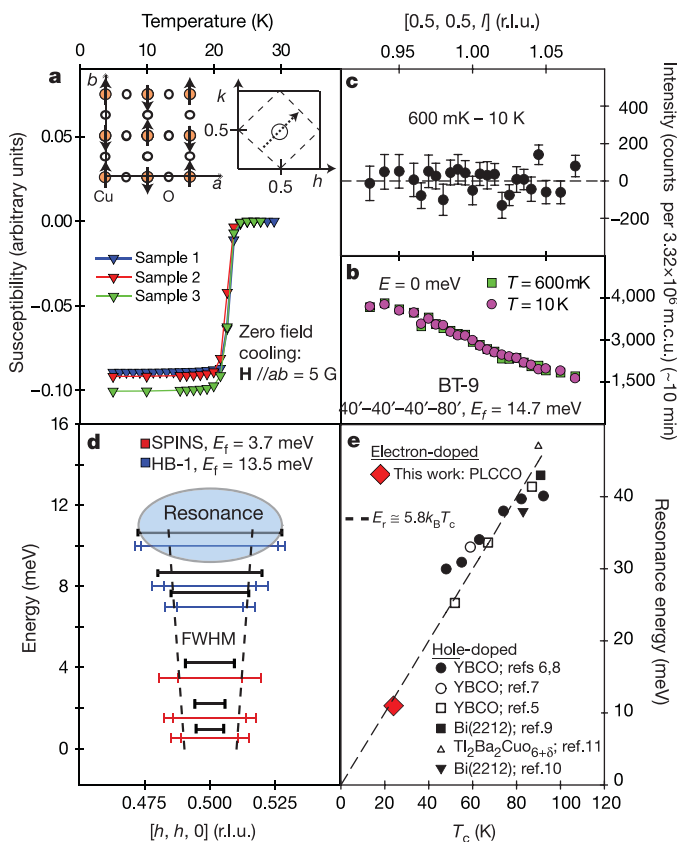
<sup>4</sup>Department of Materials Science and Engineering, University of Maryland, College Park, Maryland 20742, USA.



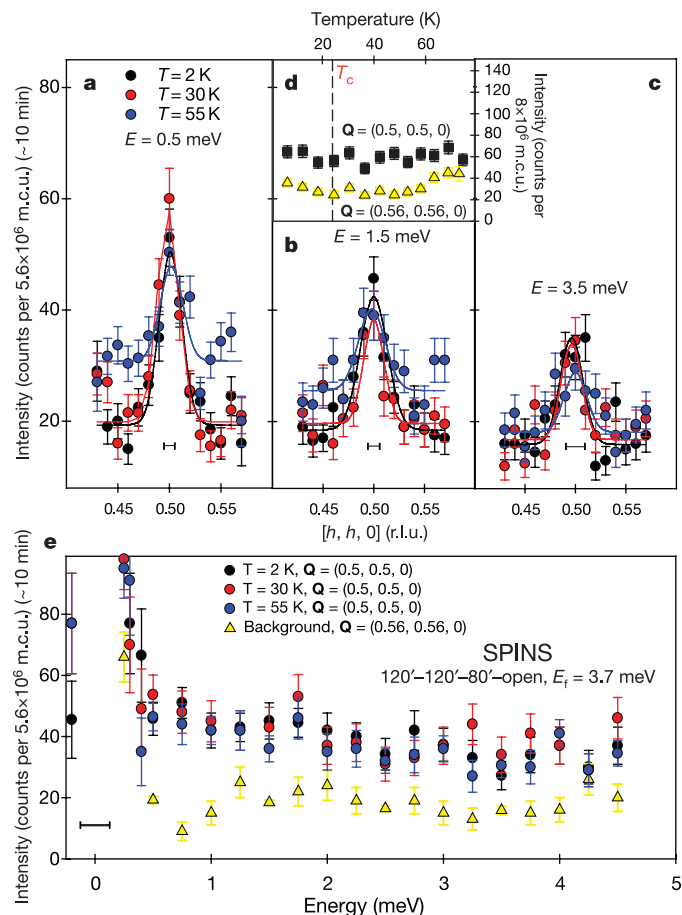
produces a shift in the background scattering on which the sharply localized magnetic excitations at  $\mathbf{Q} = (1/2, 1/2, 0)$  rest (see Supplementary Information). The temperature difference spectrum (2 K–30 K) shows a clear intensity gain around  $\sim 11$  meV at  $\mathbf{Q} = (1/2, 1/2, 0)$  (Fig. 3e).

Figure 4a shows the energy dependence of the scattering at  $\mathbf{Q} = (1/2, 1/2, 0)$  and background (0.6, 0.6, 0) positions using BT-9. The scattering at  $(1/2, 1/2, 0)$  around  $\sim 11$  meV is systematically higher below  $T_c$  while the background intensity at (0.6, 0.6, 0) is temperature-independent between 2 K and 30 K. The temperature difference spectrum below and above  $T_c$  (2 K–30 K) in Fig. 4c shows a clear

resolution-limited resonance peak centred at  $\hbar\omega \approx 11$  meV, consistent with Fig. 3e. A constant energy scan at  $\hbar\omega = 10$  meV confirms that the peak is centred at  $(1/2, 1/2, 0)$  (Fig. 4b) while a similar scan at  $\hbar\omega = 15$  meV shows only background scattering (Fig. 4b). Finally, in Fig. 4d we plot the temperature dependence of the scattering at  $(1/2, 1/2, 0)$  and the integrated intensity around  $(1/2, 1/2, 0)$  above the background for  $\hbar\omega = 10$  meV. They both increase dramatically below the onset of  $T_c$  and are remarkably similar to that of the optimally doped YBCO<sup>3,4</sup> and Bi(2212)<sup>9</sup>.

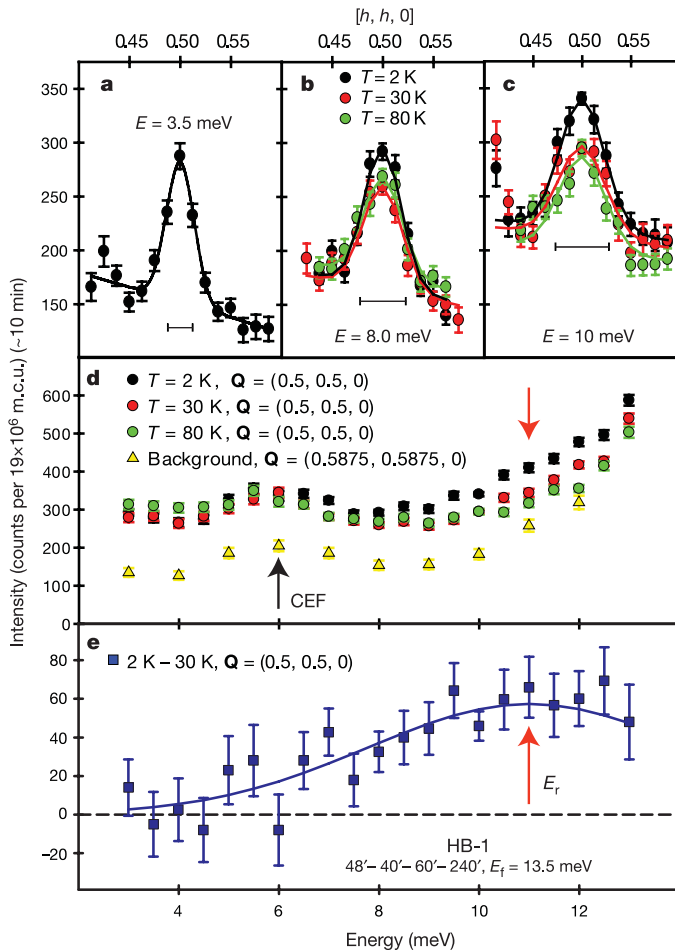


**Figure 1 | Magnetic susceptibility and a summary of neutron-scattering results.** **a**, Schematic diagrams of real and reciprocal space of the  $\text{CuO}_2$  with the dashed box showing the first Brillouin zone and the dotted arrow indicating the Q-scan direction. The temperature dependence of the magnetic susceptibility for the three crystals (mosaicity  $< 1^\circ$ ) was investigated. Our neutron-scattering experiments were performed using pyrolytic graphite for monochromators and analysers on the HB-1 triple-axis spectrometer at the High-Flux Isotope Reactor, Oak Ridge National Laboratory, and on the SPINS and BT-9 triple-axis spectrometers at the NIST Center for Neutron Research. We label the momentum transfer  $\mathbf{Q} = (q_x, q_y, q_z)$  as  $(h, k, l) = (q_x/2\pi, q_y/2\pi, q_z/2\pi)$  in the reciprocal lattice units (r.l.u.) appropriate for the tetragonal unit cell of PLCCO ( $a = b = 3.98$ , and  $c = 12.27$  Å). The three single crystals, with a total mass of  $\sim 9$  g, are co-aligned to within  $1^\circ$  in the  $[h, h, l]$  or  $[h, k, 0]$  zones. **b**, Elastic scattering along the  $[0.5, 0.5, l]$  direction through the  $(0.5, 0.5, 1)$  antiferromagnetic Bragg position at 600 mK and 10 K (ref. 23). The temperature difference spectrum in **c** shows no static antiferromagnetic order at 600 mK. In all figures, the vertical error bars are statistical uncertainties ( $1\sigma$ ) assuming a Poisson distribution function as defined in intensity-type measurements. **d**, The full-width at half-maximum (FWHM) of the magnetic response with the instrumental resolution marked as horizontal bars. The dashed lines are guides to the eye. **e**, Summary of the resonance energy as a function of  $T_c$  for hole-doped YBCO (refs 3–8), Bi(2212) (refs 9, 10),  $\text{Ti}_2\text{Ba}_2\text{CuO}_{6+\delta}$  (ref. 11), and electron-doped PLCCO (this work). The dashed line is the best fit with  $E_r = 5.8 k_B T_c$ . It takes about 10 min to run the specified monitor count units (m.c.u.).



**Figure 2 | The wavevector, energy and temperature dependence of the magnetic scattering around  $\mathbf{Q} = (1/2, 1/2, 0)$  for  $0.5 \leq \hbar\omega \leq 4.5$  meV.** The experiments were performed on SPINS with a fixed neutron final energy  $E_f = 3.7$  meV and a cold Be filter before the analyser. Error bars are  $1\sigma$ . **a–c**, Q-scans along the  $[h, h, 0]$  direction for  $\hbar\omega = 0.5, 1.5$  and  $3.5$  meV at  $T = 2, 30$ , and  $55$  K. Centred brackets (in all figures) indicate instrumental resolutions. For  $\hbar\omega = 0.5$  meV, gaussian fits on linear backgrounds give an amplitude  $A = 31.3 \pm 4.2$  counts per 10 min, width  $W = 0.011 \pm 0.0016$  r.l.u., background =  $19.2 \pm 1.2$  counts per 10 min at 2 K;  $A = 38.3 \pm 4.9$  counts per 10 min,  $W = 0.0104 \pm 0.0014$  r.l.u., background =  $19.9 \pm 1.2$  counts per 10 min at 30 K. For  $\hbar\omega = 3.5$  meV,  $A = 19.9 \pm 5.2$  counts per 10 min,  $W = 0.0098 \pm 0.0027$  r.l.u., background =  $16.1 \pm 1.3$  counts per 10 min at 2 K;  $A = 16.6 \pm 4.9$  counts per 10 min,  $W = 0.0094 \pm 0.0029$  r.l.u., background =  $16.9 \pm 1.2$  counts per 10 min at 30 K. At  $\hbar\omega = 0.5$  meV, the integrated intensities (defined as the sum of raw scanned intensities above the linear fitted background) are  $96 \pm 15$  counts per 10 min at 2 K and  $105 \pm 14$  counts per 10 min at 30 K. At  $\hbar\omega = 3.5$  meV, they are  $61 \pm 12$  counts per 10 min at 2 K and  $61 \pm 12$  counts per 10 min at 30 K. **d**, Temperature-dependent scattering at  $\mathbf{Q} = (1/2, 1/2, 0)$  and  $(0.56, 0.56, 0)$  for  $\hbar\omega = 1.5$  meV. The background is independent of temperature below  $\sim 50$  K and the magnetic signal becomes much weaker at 80 K. **e**, Constant-Q scans obtained on SPINS with spectrometer collimations as shown, at the ridge of magnetic scattering at 2, 30 and 55 K, compared with the temperature-independent background scattering for  $2 \leq T \leq 30$  K.

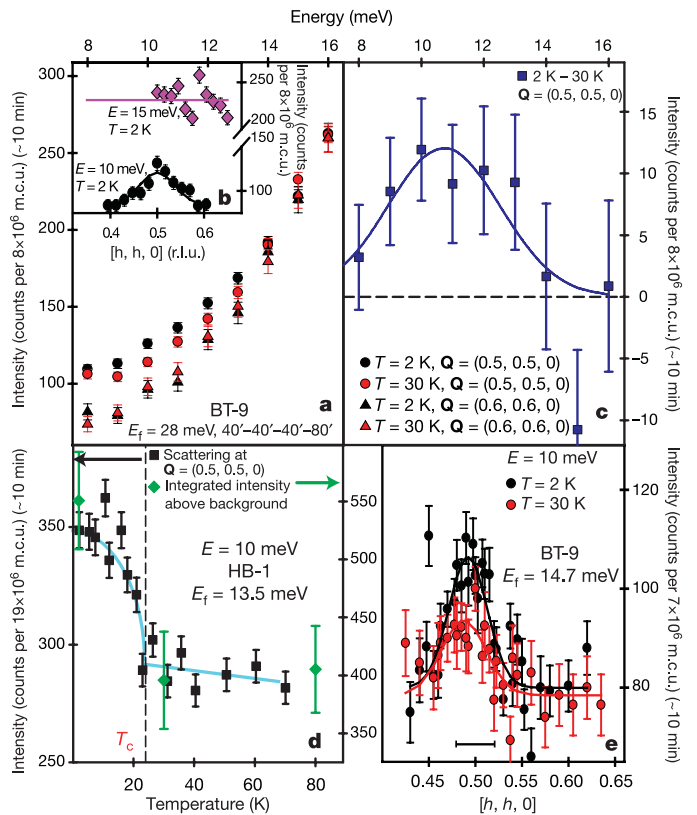
To summarize the neutron scattering results in Figs 2–4, we plotted the dispersion of the observed magnetic excitations in Fig. 1d and the resonance energy as a function of  $T_c$  in Fig. 1e (refs 3–12). Although the magnetic excitations are broader than the instrumental resolution for  $\hbar\omega \leq 3.5$  meV and resolution-limited for  $\hbar\omega \geq 4.5$  meV, they are commensurate and centred at  $(1/2, 1/2, 0)$  at all measured energies. This differs from the hole-doped materials, where incommensurate spin fluctuations below the resonance merge into it<sup>18–21</sup>. On the other hand, the resonance energy of 11 meV ( $E_r \approx 5.3k_B T_c$ )



**Figure 3 | The wavevector and energy dependence of the scattering around  $Q = (1/2, 1/2, 0)$  below and above  $T_c$ .** The experiments were carried out on HB-1 with  $E_f = 13.5$  meV and a pyrolytic graphite filter before the analyser. Error bars are  $1\sigma$ . **a–c**, Q-scans along the  $[h, h, 0]$  direction for  $\hbar\omega = 3.5, 8$  and  $10$  meV at  $T = 2, 30$  and  $80$  K. The gradual rise in the background scattering with increasing energy (and decreasing scattering angle) is due to the tail of the strong  $\text{Pr}^{3+}$  crystalline electric-field excitation at  $18$  meV and the low-angle background scattering<sup>27</sup>. For  $\hbar\omega = 10$  meV, gaussian fits on sloped linear backgrounds give amplitudes  $A = 122 \pm 4$  counts per  $10$  min, width  $W = 0.023 \pm 0.001$  r.l.u., background =  $218 \pm 9$  counts per  $10$  min at  $2$  K;  $A = 84 \pm 9$  counts per  $10$  min,  $W = 0.024 \pm 0.004$  r.l.u., background =  $212 \pm 7$  counts per  $10$  min at  $30$  K; and  $A = 87 \pm 9$  counts per  $10$  min,  $W = 0.023 \pm 0.004$  r.l.u., background =  $200 \pm 7$  counts per  $10$  min at  $80$  K. **d**, Energy scans at  $Q = (1/2, 1/2, 0)$  where the magnetic scattering peaks, and background scattering at  $(0.5875, 0.5875, 0)$ . The maximum energy transfer of  $13$  meV at  $Q = (1/2, 1/2, 0)$  is limited by kinematic constraints. The weak dispersionless peak at  $6$  meV arises from a previously unknown  $\text{Pr}^{3+}$  crystalline electric-field (CEF) level. Its intensity is temperature-independent between  $2$  and  $30$  K within the statistics of our measurements and thus can be regarded as background scattering. **e**, The temperature difference spectrum between  $2$  and  $30$  K suggests a resonance-like enhancement at  $\sim 11$  meV. See the Supplementary Information for more details on the temperature dependence of the crystalline electric-field levels.

for the PLCCO is remarkably close to the universal value of  $E_r \approx 5.8k_B T_c$  for all materials (Fig. 1e)<sup>3–12</sup>.

Our results reveal several important conclusions for the electron-hole symmetry of the magnetic excitations in high- $T_c$  copper oxides. First, the discovery of the magnetic resonance in electron-doped PLCCO with  $E_r \approx 5.3k_B T_c$  suggests that the resonance is a common feature for high- $T_c$  superconductors irrespective of electron- or hole-doping. Second, the observation of commensurate spin fluctuations below the resonance (Fig. 1d) implies that the intimate connection between incommensurate spin fluctuations and the resonance in hole-doped materials is not a universal feature<sup>18–21</sup>. At present, it is unclear how stripe models can account for these differences<sup>20,28</sup>. Third, the magnetic excitations ( $0.5 \leq \hbar\omega \leq 16$  meV) in the electron-doped PLCCO are gapless, decrease monotonically with increasing



**Figure 4 | The wavevector, energy and temperature dependence of the scattering around  $Q = (1/2, 1/2, 0)$ .** Data in **a–c** are collected using BT-9 with  $E_f = 28$  meV and a pyrolytic graphite filter before the analyser. This geometry allows the kinematic constraints to be satisfied at  $(1/2, 1/2, 0)$  for  $\hbar\omega \leq 16$  meV. Error bars are  $1\sigma$ . **a**, The energy scans along the ridge of magnetic scattering at  $(1/2, 1/2, 0)$  were counted for  $\sim 2$  h per point to obtain the statistics shown. The background scattering at  $(0.6, 0.6, 0)$  was counted for  $\sim 30$  min per point and showed no observable difference between  $2$  K and  $30$  K. **b**, Wavevector scans along the  $[h, h, 0]$  direction around  $(1/2, 1/2, 0)$  at  $\hbar\omega = 10$  and  $15$  meV. The kinematic constraints allow only half of the Q-scan at  $\hbar\omega = 15$  meV. **c**, Temperature difference ( $2$  K– $30$  K) spectrum at  $Q = (1/2, 1/2, 0)$  shows the resolution-limited resonance at  $\hbar\omega = 11$  meV. The energy resolution of the spectrometer is  $\sim 3.7$  meV in FWHM at  $\hbar\omega = 10$  meV. **d**, Black squares show temperature dependence of the neutron intensity ( $\sim 1$  h per point) at  $(1/2, 1/2, 0)$  and  $10$  meV obtained on HB-1 (Fig. 3). Green diamonds are integrated intensity of the localized signal centred around  $Q = (1/2, 1/2, 0)$  above background in Fig. 3c. The line is a guide to the eye. **e**, Q-scans at  $\hbar\omega = 10$  meV obtained on BT-9 with  $E_f = 14.7$  meV and collimations  $40^\circ$ – $40^\circ$ – $40^\circ$ – $80^\circ$ . The gaussian fits have  $A = 25.4 \pm 3.4$  counts per  $10$  min,  $W = 0.022 \pm 0.004$  r.l.u., background =  $80 \pm 3$  counts per  $10$  min at  $2$  K and  $A = 15.2 \pm 2.7$  counts per  $10$  min,  $W = 0.023 \pm 0.005$  r.l.u., background =  $78 \pm 2$  counts per  $10$  min at  $30$  K.

energy, and are virtually temperature-independent between  $2\text{ K} \leq T \leq 30\text{ K}$  except for the appearance of the resonance below  $T_c$  (Figs 2–4). Such behaviour does differ from optimally hole-doped  $\text{YBCO}^{3-6}$  and  $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$  (ref. 21), but the temperature-independent low-energy ( $0.5\text{ meV} \leq \hbar\omega \leq 4.5\text{ meV}$ ) magnetic scattering is remarkably similar to the quantum critical scattering in heavy fermion  $\text{UCu}_{5-x}\text{Pd}_x$  (ref. 29). Finally, the discovery of the magnetic resonance in electron-doped PLCCO, in which charged quasiparticles can also be probed by photoemission<sup>30</sup>, allows a systematic comparison of their properties in the same bulk sample, which has not hitherto been possible for any other high- $T_c$  superconductors. This should open new avenues of research aiming to understand the exotic properties of high- $T_c$  copper oxides.

Received 9 February; accepted 27 April 2006.

- Bardeen, J., Cooper, L. N. & Schrieffer, J. R. Theory of superconductivity. *Phys. Rev.* **108**, 1175–1204 (1957).
- Orenstein, J. & Millis, A. J. Advances in the physics of high-temperature superconductivity. *Science* **288**, 468–474 (2000).
- Rossat-Mignod, J. *et al.* Neutron scattering study of the  $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$  system. *Physica C* **185**, 86–92 (1991).
- Mook, H. A. *et al.* Polarized neutron determination of the magnetic excitations in  $\text{YBa}_2\text{Cu}_3\text{O}_7$ . *Phys. Rev. Lett.* **70**, 3490–3493 (1993).
- Fong, H. F. *et al.* Spin susceptibility in underdoped  $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ . *Phys. Rev. B* **61**, 14773–14786 (2000).
- Dai, P., Mook, H. A., Hunt, R. D. & Doğan, F. Evolution of the resonance and incommensurate spin fluctuations in superconducting  $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ . *Phys. Rev. B* **63**, 054525 (2001).
- Stock, C. *et al.* Dynamic stripes and resonance in the superconducting and normal phases of  $\text{YBa}_2\text{Cu}_3\text{O}_{6.5}$  ortho-II superconductor. *Phys. Rev. B* **69**, 014502 (2004).
- Hayden, S. M., Mook, H. A., Dai, P., Perring, T. G. & Doğan, F. The structure of the high-energy spin excitations in a high-transition-temperature superconductor. *Nature* **429**, 531–534 (2004).
- Fong, H. F. *et al.* Neutron scattering from magnetic excitations in  $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ . *Nature* **398**, 588–591 (1999).
- He, H. *et al.* Resonant spin excitations in an overdoped high temperature superconductor. *Phys. Rev. Lett.* **86**, 1610–1613 (2001).
- He, H. *et al.* Magnetic resonant mode in the single-layer high-temperature superconductor  $\text{Tl}_2\text{Ba}_2\text{CuO}_{6+\delta}$ . *Science* **295**, 1045–1047 (2002).
- Dai, P. *et al.* The magnetic excitation spectrum and thermodynamics of high- $T_c$  superconductors. *Science* **284**, 1344–1347 (1999).
- Damascelli, A., Hussain, Z. & Shen, Z.-X. Angle-resolved photoemission studies of the cuprate superconductors. *Rev. Mod. Phys.* **75**, 473–541 (2003).
- Norman, M. R. & Pepin, C. The electron nature of high temperature cuprate superconductors. *Rep. Prog. Phys.* **66**, 1547–1610 (2003).
- Basov, D. N. & Timusk, T. Electrodynamics of high- $T_c$  superconductors. *Rev. Mod. Phys.* **77**, 721–779 (2005).
- Zasadzinski, J. F. *et al.* Correlation of tunneling spectra in  $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$  with the resonance spin excitation. *Phys. Rev. Lett.* **87**, 067005 (2001).
- Hwang, J., Timusk, T. & Gu, G. D. High-transition-temperature superconductivity in the absence of the magnetic-resonance mode. *Nature* **427**, 714–717 (2004).
- Arai, M. *et al.* Incommensurate spin dynamics of underdoped superconductor  $\text{YBa}_2\text{Cu}_3\text{O}_{6.7}$ . *Phys. Rev. Lett.* **83**, 608–611 (1999).
- Bourges, P. *et al.* The spin excitation spectrum in superconducting  $\text{YBa}_2\text{Cu}_3\text{O}_{6.85}$ . *Science* **288**, 1234–1237 (2000).
- Tranquada, J. M. *et al.* Quantum magnetic excitations from stripes in copper oxide superconductors. *Nature* **429**, 534–538 (2004).
- Christensen, N. B. *et al.* Dispersive excitations in the high-temperature superconductor  $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ . *Phys. Rev. Lett.* **93**, 147002 (2004).
- Tranquada, J. M. Neutron scattering studies of antiferromagnetic correlations in cuprates. Preprint at (<http://arXiv.org/cond-mat/0512115>) (2005).
- Dai, P. *et al.* Electronic inhomogeneity and competing phases in electron-doped superconducting  $\text{Pr}_{0.88}\text{LaCe}_{0.12}\text{CuO}_{4-\delta}$ . *Phys. Rev. B* **71**, 100502(R) (2005).
- Kang, H. J. *et al.* Electronically competing phases and their magnetic field dependence in electron-doped nonsuperconducting and superconducting  $\text{Pr}_{0.88}\text{LaCe}_{0.12}\text{CuO}_{4-\delta}$ . *Phys. Rev. B* **71**, 214512 (2005).
- Yamada, K. *et al.* Commensurate spin dynamics in the superconducting state of an electron-doped cuprate superconductor. *Phys. Rev. Lett.* **90**, 137004 (2003).
- Fujita, M. *et al.* Magnetic and superconducting phase diagram of electron-doped  $\text{Pr}_{1-x}\text{LaCe}_x\text{CuO}_4$ . *Phys. Rev. B* **67**, 014514 (2003).
- Boothroyd, A. T., Doyle, S. M., Paul, McK. D. & Osborn, R. Crystal-field excitations in  $\text{Nd}_2\text{CuO}_4$ ,  $\text{Pr}_2\text{CuO}_4$ , and related  $n$ -type superconductors. *Phys. Rev. B* **45**, 10075–10086 (1992).
- Kivelson, S. A. *et al.* How to detect fluctuating stripes in the high-temperature superconductors. *Rev. Mod. Phys.* **75**, 1201–1241 (2003).
- Aronson, M. C. *et al.* Non-Fermi-liquid scaling of the magnetic response in  $\text{UCu}_{5-x}\text{Pd}_x$  ( $x = 1, 1.5$ ). *Phys. Rev. Lett.* **75**, 725–728 (1995).
- Matsui, H. *et al.* Direct observation of a nonmonotonic  $d_{x^2-y^2}$ -wave superconducting gap in the electron-doped high- $T_c$  superconductor  $\text{Pr}_{0.89}\text{LaCe}_{0.11}\text{CuO}_4$ . *Phys. Rev. Lett.* **95**, 017003 (2005).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank E. Dagotto, H. Ding and S. Zhang for discussions. We also thank Y. Ando's group for teaching us how to grow high-quality single crystals of PLCCO. S.D.W. and S.L. are supported by the US National Science Foundation. S.C. is supported by the US DOE Division of Materials Science, Basic Energy Sciences. Oak Ridge National Laboratory is supported by the US DOE through UT/Battelle LLC. SPINS is supported by the US National Science Foundation through the Center for High Resolution Neutron Spectroscopy.

**Author Information** Reprints and permissions information is available at [npg.nature.com/reprintsandpermissions](http://npg.nature.com/reprintsandpermissions). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.D. ([daip@ornl.gov](mailto:daip@ornl.gov)).



# Three-dimensional mapping of a deformation field inside a nanocrystal

Mark A. Pfeifer<sup>1†</sup>, Garth J. Williams<sup>1†</sup>, Ivan A. Vartanyants<sup>1†</sup>, Ross Harder<sup>1</sup> & Ian K. Robinson<sup>1†</sup>

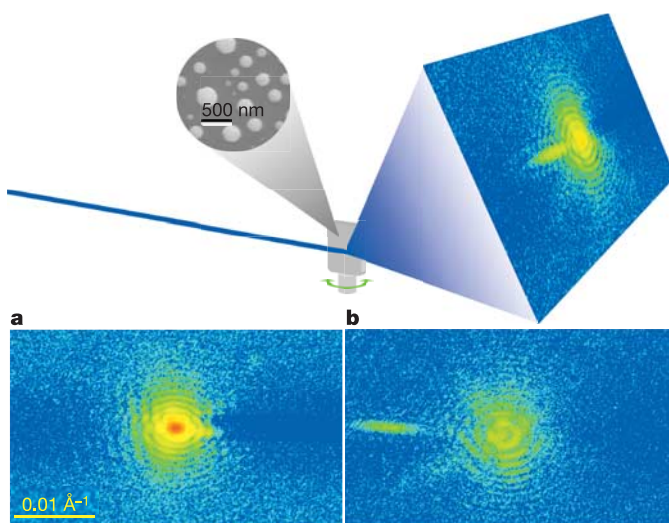
Coherent X-ray diffraction imaging is a rapidly advancing form of microscopy: diffraction patterns, measured using the latest third-generation synchrotron radiation sources, can be inverted to obtain full three-dimensional images of the interior density within nanocrystals<sup>1–3</sup>. Diffraction from an ideal crystal lattice results in an identical copy of this continuous diffraction pattern at every Bragg peak. This symmetry is broken by the presence of strain fields, which arise from the epitaxial contact forces that are inevitable whenever nanocrystals are prepared on a substrate<sup>4</sup>. When strain is present, the diffraction copies at different Bragg peaks are no longer identical and contain additional information, appearing as broken local inversion symmetry about each Bragg point. Here we show that one such pattern can nevertheless be inverted to obtain a ‘complex’ crystal density, whose phase encodes a projection of the lattice deformation. A lead nanocrystal was crystallized in ultrahigh vacuum from a droplet on a silica substrate and equilibrated close to its melting point. A three-dimensional image of the density, obtained by inversion of the coherent X-ray diffraction, shows the expected faceted morphology, but in addition reveals a real-space phase that is consistent with the three-dimensional evolution of a deformation field arising from interfacial contact forces. Quantitative three-dimensional imaging of lattice strain on the nanometre scale will have profound consequences for our fundamental understanding of grain interactions and defects in crystalline materials<sup>4</sup>. Our method of measuring and inverting diffraction patterns from nanocrystals represents a vital step towards the ultimate goal of atomic resolution single-molecule imaging that is a prominent justification for development of X-ray free-electron lasers<sup>5–7</sup>.

Coherent X-ray diffraction imaging emerged from the realization that oversampled diffraction patterns can be inverted to obtain real-space images. The possibility was first pointed out by Sayre<sup>1</sup> but not demonstrated until 1999 by Miao *et al.*<sup>2</sup>. The phase information of the diffraction pattern, which is lost in its recording, is embedded in a sufficiently oversampled diffraction pattern because this is intimately related to the Fourier transform of the object under investigation. The inversion of diffraction back to an image has been proven to be unique in two or higher dimensions, except for ‘pathological’ cases of internal symmetry of the object or its diffraction pattern<sup>8,9</sup>. Computational methods of performing the inversion are being actively developed; they are often based on the iterative ‘hybrid input–output’ (HIO) method introduced in the 1980s by Fienup<sup>10</sup>.

Lensless imaging using coherent X-rays is an attractive alternative to electron microscopy because of the better penetration of the electromagnetic waves in materials of interest; multiple scattering effects can be neglected so that the first Born approximation can safely be used.

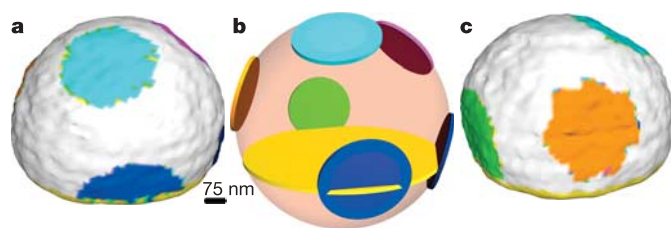
In some cases X-rays are less damaging to the sample than electrons; also, the collection of a diffraction pattern is inherently more efficient than the use of lenses<sup>11</sup>. If the diffraction can be reliably inverted by computation, the method could be used routinely to reveal the structure of materials on the nanometre scale, far beyond the resolution of the traditional light microscope. The holographic method of combining a reference wave is an alternative way to perform the inversion<sup>12</sup>.

The use of short-wavelength X-rays and crystalline materials introduces the additional possibility of Bragg diffraction. In the



**Figure 1 | Schematic diagram of the coherent X-ray diffraction experiment.** Pb was evaporated onto a heated Si wafer support, with its native oxide intact, to make a film of about 20 nm thickness. After melting the film, it formed molten droplets, which were then cooled to overcome the (substantial) supercooling until the liquid crystallized, then raised again to 1.2 K below the Pb melting point of 600.6 K. Ultrahigh vacuum conditions were maintained throughout. Later examination by SEM (circular inset) showed isolated hemispherical crystals. Undulator X-rays from the Advanced Photon Source (APS) were monochromated using Si(111), selecting a wavelength of 1.38 Å, and collimated by narrow slits to illuminate a few hundred crystals of the sample. A direct-reading CCD X-ray detector, 1.32 m away, was centred on the (111) Bragg peak of one of the crystals, to give the diffraction patterns shown. A rotation series of 50 diffraction patterns was collected by rotating the sample in steps of 0.01° about the axis shown. Two representative frames are shown in **a** and **b**, while a fuller series is given in Supplementary Fig. S2. The total exposure time of each frame was 150 s.

<sup>1</sup>Physics Department, University of Illinois, Urbana, Illinois 61801, USA. <sup>†</sup>Present addresses: Department of Physics, University of Oregon, Eugene, Oregon 97403, USA (M.A.P.); School of Physics, University of Melbourne, Melbourne, Victoria, Australia (G.J.W.); HASYLAB, DESY, Hamburg, Germany (I.A.V.); Department of Physics and Astronomy, University College London, London, UK (I.K.R.).



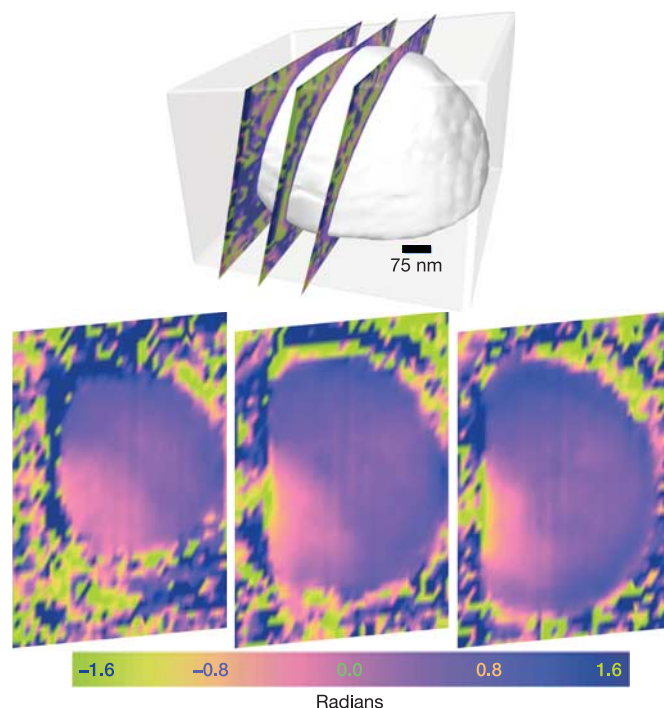
**Figure 2 | Two perpendicular views of the three-dimensional reconstruction of the magnitude of the nanocrystal's complex density function.** The density function is shown as 50% density isosurfaces in **a** and **c**. The fitted facet planes of the equilibrium crystal shape have been coloured. **b**, Schematic model of a sphere (semi-transparent) and facet planes (disks) fitting the view of **a**. The diameter of the nanocrystal is 750 nm.

simplest approximation of ideal crystals, this yields an identical copy of the forward diffraction centred around each Bragg peak. Using Bragg diffraction patterns not only allows individual grains to be selected for analysis one by one and avoids losing data behind a beam-stop<sup>2</sup>, but also greatly facilitates the recording in three dimensions, as demonstrated for micrometre-sized gold crystals<sup>3</sup>. The inversion is formally identical to that of the forward scattering<sup>2</sup>, but reveals only the density of the crystalline part of the sample, and so is highly sensitive to defects.

Diffraction also opens the new possibility of directly imaging the strain fields within the crystal, an opportunity that we exploit in this work. It is easy to demonstrate that the presence of strain breaks the local symmetry of a diffraction pattern about the Bragg point, which would otherwise show inversion symmetry (as it does about the origin, according to Friedel's law). Whereas the inverse Fourier transform of a symmetric function is real, that of the asymmetric distribution must in general be complex, possessing phase structure in real space. It has been shown<sup>13</sup> that, without loss of generality, the density of a crystal can be considered to be a complex function whose magnitude is the physical electron density and whose phase is the projection of the local deformations of the crystal lattice onto the reciprocal lattice vector  $\mathbf{Q}$  of the Bragg peak about which the diffraction is measured.

Because there are twice as many independent measurement points for an asymmetric diffraction pattern than a centrosymmetric one and twice as many variables needed to describe a complex density function as a real one, the problem is overdetermined to the same degree. The oversampling ratio<sup>14</sup>, which determines whether such a pattern can be inverted, is the same. The primary measure of the oversampling of the diffraction is the size of the 'support' region that is the volume of space within which the finite-sized object is confined to exist<sup>14</sup>; this constraint is unchanged for the complex problem. In our experience with test calculations, we have found no additional difficulty in the convergence of the HIO-like algorithms for the complex problem, and this has been confirmed by others<sup>15,16</sup>. The greatest sensitivity arises from the choice of the support constraint, just as it does for the real density problem.

After our earlier preparation of free-standing Au nanocrystals<sup>3</sup>, we developed a method of growing nanometre-sized Pb crystals inside the vacuum chamber of the 34-ID-C beamline at the Advanced Photon Source (APS), shown schematically in Fig. 1. X-ray diffraction measurements were then made *in situ* of the continuous diffraction pattern surrounding a (111) Bragg peak. Figure 1 and Supplementary Fig. S2 show slices through the concentric shells of intensity that correspond to the shape transform of the hemispherical crystal. Additionally, a number of flares can be seen, which arise from facets formed during solidification of the crystal. Scanning electron microscope (SEM) images, also shown in Fig. 1, subsequently found the sample to contain an array of isolated crystals with a faceted

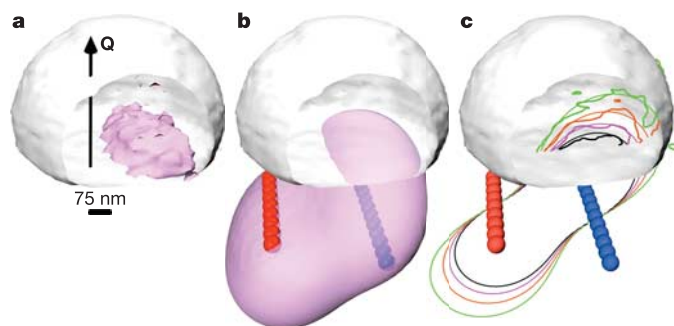


**Figure 3 | Phase maps cutting through the crystal at three parallel planes.** Schematic diagram of the sections, 138 nm apart. The translucent box is the support region used in the phasing calculations, which was rectangular before the coordinate transformation. The phase bulge is interpreted as a projection of strain fields in the crystal lattice arising from contact forces at the interface with the substrate. Further sections of the phase map are provided in Supplementary Fig. S5.

hemispherical morphology and a distribution of sizes centred around 200 nm.

The three-dimensional data were inverted to real space in the coordinate system of the measurement, following the two directions of the charge-coupled device (CCD) camera's pixel array and the sample rotation angle. Initial attempts to phase the background-subtracted diffraction pattern were made using established methods<sup>3,10,14–17</sup>. We applied as many as seven alternating cycles of 'error reduction' (ER) and the HIO algorithm, each consisting of many iterations of the respective algorithm<sup>10</sup>, with real-valuedness and finite support as the real-space constraints. The support region was allowed to shrink by thresholding the images obtained by prior inversions with a larger support<sup>17</sup>. Despite using a tight support, the recovered density included a large region of low density within the crystal boundary. This region featured a non-uniform phase despite constraining the phases to be zero on every iteration.

When we phased the diffraction pattern without enforcing a real-space phase constraint, we obtained qualitatively better results. Fifteen phasing calculations were performed in a  $324 \times 192 \times 48$  array, comprising 200 iterations of HIO followed by 1,000 iterations of ER, each starting with different random starting phases, using a simple  $41 \times 37 \times 24$  rectangular support with more than twice the volume of the tight support used previously. All fifteen results had an error metric, defined as the normalized integrated squared difference between the Fourier transform of the image and the data, between 0.022 and 0.023. When the results with the three lowest error metrics were compared, they showed a reproducibility of 0.024 (analogous definition), meaning that, within the accuracy of the data, they were identical<sup>18</sup>. The magnitude of the complex density (Fig. 2 and Supplementary Fig. S4) is then assumed to be the physical density, while the phase (Fig. 3 and Supplementary Fig. S5) is interpreted as the scalar product of the local deformation and the



**Figure 4 | Phase interpretation.** Single isosurface of the reconstructed phase of the complex density function (**a**) and its best fit (**b**), superimposed on a cut-away image of the crystal density. The point defect lines used to generate the fit (dots) and the direction of  $Q$  (arrow) are also illustrated. An animated version of this figure is presented as Supplementary Fig. S1. **c**, Contour map of the reconstructed phase on a cross-section plane passing near the middle of the nanocrystal in the same view. Smooth lines are the corresponding contours of the projection function  $Q \cdot u(r)$  where  $u(r)$  is the strain field calculated for two rows of point defects (balls) of opposite sign. Both sets of contours have spacings of 0.24 radians.

momentum transfer vector<sup>13</sup>  $Q$ . A geometric calculation was then used to transform the result to orthogonal directions in real space<sup>19</sup> for viewing the image in a cartesian laboratory frame, as shown in Supplementary Fig. S3. These images showed a relatively uniform internal density, varying by about 15%, with a phase gradient where the density had been weaker before.

The physical density map in Supplementary Fig. S4 showed the crystal shape to be a sphere of diameter 750 nm with flat facets, as expected<sup>20</sup> for the equilibrium crystal shape of Pb, owing to the anisotropy of its surface free energy<sup>21</sup>  $\gamma$ . Angular variations of  $\gamma$  cause faceting along certain crystallographic directions through the Wulff construction<sup>21</sup>. A least-squares fit to the density boundary of our sample, shown in Fig. 2 and Supplementary Fig. S6, identified clear facets along  $\{111\}$  directions as well as the plane of the interface with the substrate, which was non-crystallographic<sup>19</sup>. To estimate the resolution of the final image, we plotted line scans of the density projected onto the normals of these facets, as well as radial scans through the spherical regions in Supplementary Fig. S7. These illustrate how steeply the density edge of the crystal drops across the boundary of the crystal. The density step could be well described in Supplementary Fig. S7 by fitting error functions of width 40 nm, which was taken to be the image resolution.

The main focus of our attention is the internal strain fields revealed by the phase of the complex density in Fig. 3 and Supplementary Fig. S5. Most of the crystal was undeformed, with a constant phase, but significant deviations appeared in an egg-shaped bulge near the middle of its interface with the substrate, rising to a maximum of 1.1 radians at the bottom edge of the crystal where it makes contact with the substrate. This corresponds to a maximum deformation of the crystal of  $(1.1/2\pi)$  (111) lattice spacings, or 0.5 Å. The phase bulge is illustrated in Fig. 4a as an isosurface of constant phase and as contour plots of a two-dimensional slice through the phase map in Fig. 4c. Across the short direction of the bulge, the equal-phase contours are seen to have a roughly circular shape. This is consistent with the radial deformation field  $u(r)$  of a classical point defect, such as a lattice interstitial or vacancy. When such a deformation field is projected onto the  $Q$ -vector, the resulting phase forms two spheres of constant  $Q \cdot u(r)$  that touch at the location of the defect<sup>22</sup>. In our image, only part of a sphere is seen, indicating that the point defect is virtual, lying outside the crystal. To simulate the long direction of the phase bulge, we constructed a model that superimposes two lines of point defects of opposite sign. The full calculated

phase isosurface from the resulting deformation is shown in Fig. 4b. This model explains both the detailed shape of the phase contours and the inverse-square variation of the strain projection with distance, consistent with the observed spacing of the contours in Fig. 4c.

The point defects assumed to explain the data are virtual sources of the deformation field, lying outside the crystal. The physical origin of this strain is a surface deformation distributed across the interface with the  $\text{SiO}_2$  substrate. The surface distortion results in a long-range strain field within the crystal, exactly analogous to the effect of surface charges in electrostatics. It is likely that the Pb crystal had nucleated from the melt at some surface defect of the substrate that imparts a residual contact strain. The surface strain distribution is connected with the measured deformation distribution by the Poisson equation governing continuum elasticity theory.

Our result shows that physically reasonable strain fields can be imaged three-dimensionally inside crystalline materials of nanometre dimensions. The deformation field seen in our Pb nanocrystal is attributed to a surface defect on the substrate where the crystal nucleated. It is clear that, by extension of this work, all of the classic defects, the basic ingredients of materials science, could be imaged in this way. The use of a nanocrystal host limits the volume of space that has to be mapped to see an isolated defect, but also serves as the support needed for the diffraction phasing to succeed. Our method of obtaining phase contrast is complementary to the real-space strain-mapping methods that depend on the preparation of nanometre-sized beams<sup>23</sup>, and to the Fresnel-zone-plate-based methods<sup>24</sup>. Our resolution is at present 40 nm, but it can be improved substantially by the development of better detectors and optics or more powerful and more coherent X-ray sources, such as an X-ray free-electron laser<sup>6</sup>. The three-dimensional imaging of complete nanocrystals, using the inversion methods demonstrated here, can be extended to atomic resolution when such measurements become possible<sup>5,7</sup>.

Received 24 October 2005; accepted 5 May 2006.

- Sayre, D. Some Implications of a theorem due to Shannon. *Acta Crystallogr.* **5**, 843 (1952).
- Miao, J., Charalambous, P., Kirz, J. & Sayre, D. Extending the methodology of X-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens. *Nature* **400**, 342–344 (1999).
- Williams, G. J., Pfeifer, M. A., Vartanyants, I. A. & Robinson, I. K. Three-dimensional imaging of microstructure in gold nanocrystals. *Phys. Rev. Lett.* **90**, 175501 (2003).
- Cahn, R. W. Epitaxy keeps rolling along. *Nature* **375**, 363–364 (1995).
- Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. & Hajdu, J. Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature* **406**, 752–757 (2000).
- Patel, N. Shorter, brighter, better. *Nature* **415**, 110–111 (2002).
- Huldt, G., Szöke, A. & Hajdu, J. Diffraction imaging of single particles and biomolecules. *J. Struct. Biol.* **144**, 219–227 (2003).
- Bates, R. H. T. Fourier phase problems are uniquely solvable in more than one dimension. *Optik* **61**, 247–262 (1982).
- Hayes, M. H. The reconstruction of a multidimensional sequence from the phase or magnitude of its Fourier transform. *IEEE Trans. Acoust. Speech Signal Process.* **30**, 140–154 (1982).
- Fienup, J. R. Phase retrieval algorithms: a comparison. *Appl. Opt.* **21**, 2758–2769 (1982).
- Henderson, R. The potential and limitations of neutrons, electrons, and X-rays for atomic resolution microscopy of unstained biological molecules. *Q. Rev. Biophys.* **28**, 171–193 (1995).
- Eisebitt, S. et al. Lensless imaging of magnetic nanostructures by X-ray spectro-holography. *Nature* **432**, 885–888 (2004).
- Vartanyants, I. A. & Robinson, I. K. Partial coherence effects on the imaging of small crystals using coherent X-ray diffraction. *J. Phys. Condensed Matter* **13**, 10593–10611 (2001).
- Miao, J., Sayre, D. & Chapman, H. N. Phase retrieval from the magnitude of the Fourier transforms of nonperiodic objects. *J. Opt. Soc. Am. A* **15**, 1662–1669 (1998).
- Fienup, J. R. Reconstruction of a complex-valued object from the modulus of its Fourier transform using a support constraint. *J. Opt. Soc. Am. A* **4**, 118–123 (1987).



16. McBride, W., O'Leary, N. L. & Allen, L. J. Retrieval of a complex-valued object from its diffraction pattern. *Phys. Rev. Lett.* **93**, 233902 (2004).
17. Marchesini, S. *et al.* X-ray image reconstruction from a diffraction pattern alone. *Phys. Rev. B* **68**, 140101 (2003).
18. Williams, G. J. *Microscopy of Gold Microcrystals by Coherent X-ray Diffractive Imaging*. PhD thesis, Univ. Illinois (2004).
19. Pfeifer, M. A. *Structural Studies of Lead Nanocrystals using Coherent X-ray Diffraction*. PhD thesis, Univ. Illinois (2004).
20. Heyraud, J. C. & Métois, J. J. Equilibrium shape and temperature: lead on graphite. *Surf. Sci.* **128**, 334–342 (1982).
21. Wortis, M. in *Chemistry and Physics of Solid Surfaces VII* (eds Vanselow, R. & Howe, R.) 367 (Springer, Berlin, 1988).
22. Metzger, T. H. & Peisl, H. Huang diffuse x-ray scattering from lattice strains in high-concentration Ta-H alloys. *J. Phys. F* **8**, 391–402 (1978).
23. Larson, B. C., Yang, W., Ice, G. E., Budai, J. D. & Tischler, J. Z. Three-dimensional X-ray structural microscopy with submicrometre resolution. *Nature* **415**, 887–890 (2002).
24. Jacobsen, C. & Kirz, J. X-ray microscopy with synchrotron radiation. *Nature Struct. Biol.* **5**, 650–653 (1998).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This research was funded by the NSF Division of Materials Research and the EPSRC. The UNICAT facility at the APS and the APS itself are supported by the US Department of Energy, the former through the Frederick Seitz Materials Research Laboratory.

**Author Information** Reprints and permissions information is available at [npg.nature.com/reprintsandpermissions](http://npg.nature.com/reprintsandpermissions). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to I.K.R. ([i.robinson@ucl.ac.uk](mailto:i.robinson@ucl.ac.uk)).

# The effect of energy feedbacks on continental strength

Klaus Regenauer-Lieb<sup>1,2</sup>, Roberto F. Weinberg<sup>3</sup> & Gideon Rosenbaum<sup>4</sup>

The classical strength profile of continents<sup>1,2</sup> is derived from a quasi-static view of their rheological response to stress—one that does not consider dynamic interactions between brittle and ductile layers. Such interactions result in complexities of failure in the brittle–ductile transition and the need to couple energy to understand strain localization. Here we investigate continental deformation by solving the fully coupled energy, momentum and continuum equations. We show that this approach produces unexpected feedback processes, leading to a significantly weaker dynamic strength evolution. In our model, stress localization focused on the brittle–ductile transition leads to the spontaneous development of mid-crustal detachment faults immediately above the strongest crustal layer. We also find that an additional decoupling layer forms between the lower crust and mantle. Our results explain the development of decoupling layers that are observed to accommodate hundreds of kilometres of horizontal motions during continental deformation.

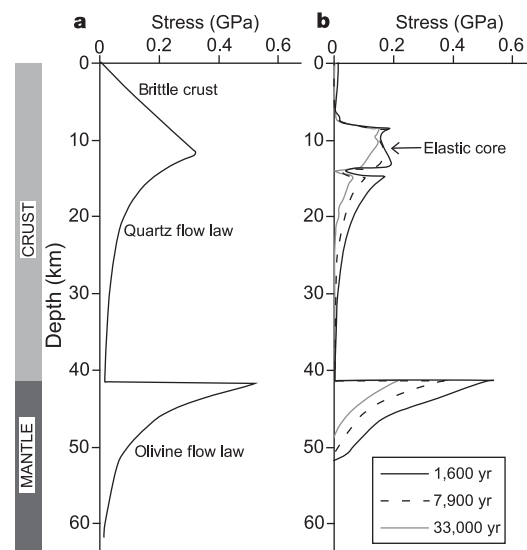
The strength profile of the continental lithosphere is understood to be a combination of the pressure-dependent brittle strength of the upper crust (Byerlee's law), and the viscous, temperature-dependent and strain-rate-dependent strength of the lower continental crust, dominated by either quartz or plagioclase rheology<sup>3</sup>. The brittle–ductile transition is typically defined for a given characteristic geological strain rate (often  $10^{-15} \text{ s}^{-1}$ ) as the depth where the two strength curves intersect. As temperature increases with depth, the lower crust becomes increasingly weaker, until reaching the mantle, which is dominated by the generally stronger olivine rheology. This is known as the Brace-Goetze strength profile<sup>1–3</sup>, or informally as the 'Christmas tree' (Fig. 1a). The weak lower crust sandwiched between the strong brittle–ductile transition area and the strong mantle defines what is known as the 'jelly sandwich'. This view is simple and powerful, and has led us a long way towards understanding the deformation of continents in the context of plate tectonics.

The 'Christmas tree' strength profile, however, implies that the strongest part of the crust is located at the layers immediately above the brittle–ductile transition (Fig. 1a). This raises a major problem in continental tectonics, because field evidence shows strain localization and the development of major detachment faults at typical depths of the brittle–ductile transition<sup>4–8</sup>. The reason for this apparent paradox may be that existing models have not considered the dynamic evolution of the strength profile due to loading and the history of deformation and strain localization.

In order to evaluate the rheological response to energy feedback effects, we designed numerical models of extension with an initial crustal thickness of 42 km. We assume a free top surface and zero tangential stress (free slip) on other boundaries. Extension is driven by velocity boundary conditions of  $1 \text{ cm yr}^{-1}$  applied on either side of the model. Implementing such boundary conditions using the

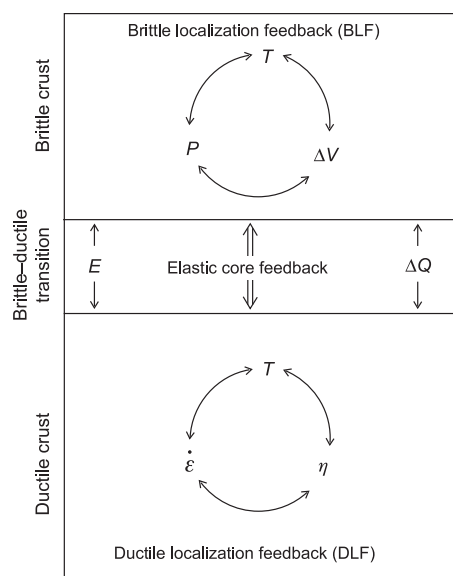
classical Mohr-Coulomb approach<sup>9</sup> produces a pure shear style of extension, which is accommodated by steeply dipping normal faults and fails to predict the development of low-angle normal faults. However, in our fully coupled simulation (see Methods and Supplementary Appendix 1), localization feedbacks develop in and between the brittle and ductile layers (Fig. 2). These are expressed by the development of listric faulting and detachment faults in the brittle upper crust and ductile shearing in an elasto-viscoplastic lower crust. In this process, brittle faulting may rupture at seismogenic rates (for example,  $10^2$ – $10^3 \text{ m s}^{-1}$ ), whereas shear zones propagate at much slower rates (up to  $3 \times 10^{-9} \text{ m s}^{-1}$ )<sup>10</sup>. This contrast in strain rates leads to complex interactions at the brittle–ductile transition<sup>11</sup>.

Complex structural evolution, related to dynamic changes in strength, is recognized in the extensional models in Fig. 3. During the early stages of loading (Fig. 3a), two distinct depth levels develop sub-horizontal high strain decoupling segments—a deeper and a shallower one. The deep decoupling zone is rooted in the lowermost



**Figure 1 | Strength profiles of the lithosphere.** **a**, Simplified Brace-Goetze strength profile for parameters in Table 1 in Supplementary Appendix 1. **b**, An example of early stress evolution for the dynamic feedback calculation, showing the development of an elastic core that is gradually eroded by the dynamic weakening in the upper mantle and the lower crust (for a more evolved stress profile, see Supplementary Appendix 3). Note that the brittle crust has failed while the elastic core retains its strength. The pronounced weak zone at  $\sim 14 \text{ km}$  is developed owing to ductile localization feedback effects.

<sup>1</sup>Institut für Geowissenschaften, Johannes Gutenberg Universität, Mainz 55099, Germany. <sup>2</sup>School of Earth and Geographical Sciences, The University of Western Australia, Perth, Western Australia 6009, Australia, and CSIRO Exploration & Mining, Kensington, Western Australia 6151, Australia. <sup>3</sup>School of Geosciences, Monash University, Clayton, Victoria 3800, Australia. <sup>4</sup>School of Physical Sciences, The University of Queensland, Brisbane, Queensland 4072, Australia.



**Figure 2 | Strain localization feedbacks in upper and lower crust.**

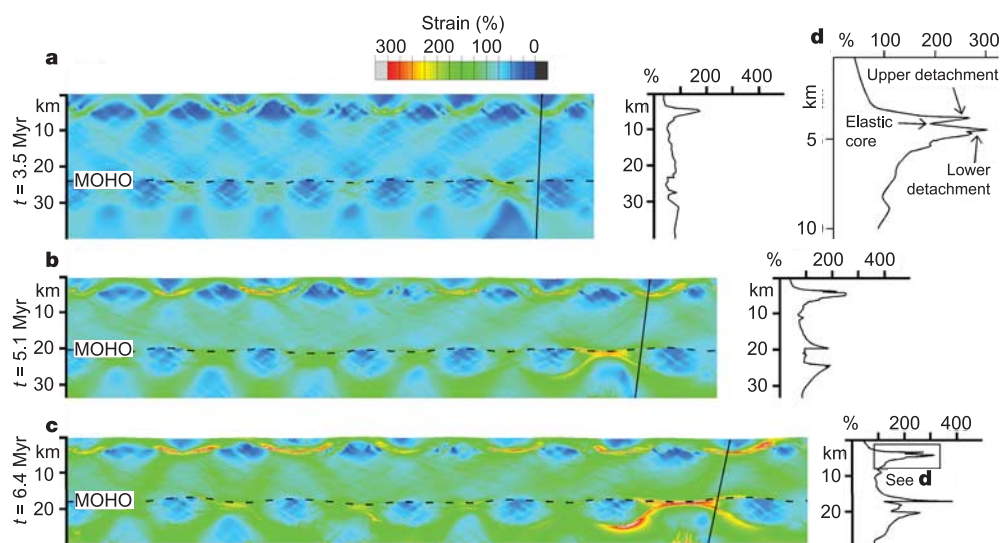
The brittle localization feedback (BLF) involves pressure ( $P$ ), temperature ( $T$ ) and volume change ( $\Delta V$ ) and gives rise to brittle faults in upper crust. The ductile localization feedback (DLF) gives rise to ductile shear zones in lower crust, and also includes components of viscosity ( $\eta$ ), strain rate ( $\dot{\epsilon}$ ) and temperature ( $T$ ), dampened by heat diffusion. The two layers communicate by changes of stored elastic energy ( $E$ ) and heat ( $\Delta Q$ ) in the strong mid-crustal elastic core, which corresponds to a brittle–ductile transition zone. The depth of this transition is a function of  $P$ ,  $T$ ,  $t$ , strain rate, shear modulus, Poisson ratio and activation energy.

crust on top of the mantle. It can be attributed to rheological contrasts between quartz and olivine. This contrast is amplified by a thermo-mechanical ductile localization feedback (DLF in Fig. 2), resulting in an effective viscosity drop just above the Moho, which increases with deformation (Supplementary Appendix 3). In the mantle, small-scale conjugate pairs of shear instabilities nucleate on a short wavelength (of the order of 100 m) around thermal pertur-

bations. These are ephemeral structures that gradually cascade to larger wavelengths. This leads to shear focusing onto a number of master shear zones (Fig. 3b). The net effect of these shear instabilities is that the mantle evolves dynamically to a weak rheology with shear stress less than 0.3 GPa (Fig. 1b). Quite opposite to the prediction of Fig. 1a, the mantle is not the strongest layer but evolves to be weaker than the upper crust (Fig. 1b and Supplementary Appendix 3).

The efficient weakening of the mantle layer relies on the development of strong shear zones reaching from the deep crust decoupling layer into the mantle. Conversely, shear zones with an upward continuation from the deep decoupling layer into shallower crust are diffuse and die out into an elastic core. This core is defined as the region of maximum elastic energy storage in the crust. It appears initially without shear zones and deforms pervasively at a slower rate than the localized shear zones in the upper and lower crusts (Supplementary Appendix 3). Above the elastic core, deformation is governed by the brittle localization feedback (BLF; Fig. 2), which forms conjugate faults (Fig. 3a). Like in the mantle, ductile shear zones in the lower crust collocate into the largest possible structures and wavelength (Fig. 3b). In the upper crust, faults gradually become listric, forming a detachment layer of maximum shear heating (that is, maximum dissipation) on top of the elastic core (Fig. 3b).

The complex deformation of the elastic core, dominated simultaneously by brittle and ductile strain localization, is the key to understanding the deformation of the lithosphere in general and the communication between the upper and lower crust in particular. This core is characterized not only by the maximum elastic energy in the crust, but also by having the lowest ratio between Young's modulus and yield strength. This leads to a competition between storing elastic energy and dissipating heat, which is translated into the development of contemporaneous short-term, rapid, brittle cracks localized on pressure anomalies, and short-term, slow, ductile shear zones on temperature anomalies. The high stored energy in this area allows many dynamic shear zones to develop. However, they die out because their size is below the critical length scale for survival, defined by the equilibration between shear heating and heat diffusion (of the order of kilometres, equivalent to the square root of the ratio between heat diffusivity and strain rate in the shear zone<sup>12</sup>).



**Figure 3 | Model results.** **a–c**, Results of the  $70 \text{ mW m}^{-2}$  heat flow extension model using parameters from Table 1 in Supplementary Appendix 1. Additional results for  $60 \text{ mW m}^{-2}$  are shown in Supplementary Appendix 2. The coloured panels show vertical cross-sections through the lithosphere plotted on a 1:1 x:y spatial scale. Note the development of mid-crustal detachments and an additional decoupling zone on top of the Moho. The

strain profile on the right of each panel is plotted for a straight line, which is initially vertical (at 0 Myr) and tilts significantly owing to differential deformation of crust and mantle, causing simple shear out of pure shear boundary conditions. **d**, Enlargement of the strain profile at 6.4 Myr, showing the development of two upper crustal detachment levels.



The propagation of the lower tip of a brittle fault is driven by the combination of loading and brittle localization feedback effects. Fault propagation is hampered as it reaches the top of the elastic core, where brittle and ductile processes interact, because it encounters a viscous response that blunts the fault tip and eliminates the stress singularity. Further arrest of the brittle fault is driven by ductile thermal-mechanical feedback effects, which act to decrease the differential stress at the fault's tip by lowering the effective viscosity<sup>11</sup>. The softening occurs at a subhorizontal layer, transforming the fault into a heat source and a detachment defined by a mylonitic shear zone.

Simultaneously, ductile shear zones propagate upwards towards the elastic core. However, these shear zones die out because of increasing crustal strength and because they reach the area in which brittle localization feedback dominates. Their energy is translated into a local, positive amplitude disturbance of the elastic stress field. This extra stress, loaded to the core from below, eventually starts to interact with similar features loaded from above. When they coalesce, the core is disrupted leading to a second, deeper detachment rooted at the base of the mid-crustal core. In this way, the upper mid-crustal detachment zone develops into a 1.5-km-wide low viscosity zone defined by two detachments separated by an elastic core (Fig. 3c, d). This detachment zone is located immediately at or near the strongest part of the crust, and coincides with a zone of low effective viscosity (Supplementary Appendix 3). Significant mantle uplift develops below the detachment, while the surface topography remains relatively moderate (570 m maximum amplitude, see Supplementary Appendix 4). The dynamic evolution of the brittle–ductile transition is controlled by three significant aspects of evolution: (1) the apparent weakness of the uppermost crust reflects the fact that stress in this layer is released on faults; (2) similarly, the gradual weakening of the uppermost mantle controls the development of collocated, wide ductile shear zones (Fig. 3b, c); and (3) development and interaction of the detachment zones above and below the narrowing elastic core, which ultimately breaks up, leading to a lithospheric scale continental break-up. Quite contrary to the quasi-static prediction of Fig. 1, it is not the mantle but the mid-crustal brittle–ductile transition which dynamically evolves into the stress-bearing part of the lithosphere for most of its deformation history.

Thus, in these models the lithosphere evolves through several steps as stress is loaded and strain is localized: first, crust and mantle decouple, followed by the weakening of the mantle through the development of shear zones. The mid-crustal detachment then develops at the top of the elastic core. The gradual weakening of the elastic core through loading then leads to the second ductile detachment at the base of the core, followed by crust-wide shear zones which lead ultimately to lithospheric break-up. In models with a lower thermal gradient, as in the  $60 \text{ mW m}^{-2}$  run, the strong elastic core is initially wider, and consequently the mid-crustal detachment has a similar but yet much richer evolution (Supplementary Appendix 2).

Crustal-scale detachments or décollements are characteristic features of thin-skinned terranes, accommodating large relative displacements between allochthonous, brittle upper slabs, and ductilely deformed lower slabs. Detachments have been documented in extensional environments (for example, Basin and Range<sup>4,5,8</sup> and the Aegean Sea<sup>13</sup>), and in shortening environments (for example, Helvetic nappes in the Alps<sup>14</sup> and Canadian Cordillera<sup>15</sup>). The extensional detachments were originally thought to penetrate the whole crust to the Moho<sup>16</sup>. However, further investigation showed that they tend to root at the brittle–ductile transition, at crustal depths around  $10 \pm 5 \text{ km}$  (refs 4–8, 17). Similarly, convergent terranes have well-established décollements at mid-crustal levels<sup>15</sup> and at the base of the crust<sup>18</sup>.

In contrast with observations, the Brace-Goetze strength profile predicts that the brittle–ductile transition is the strongest part of the crust and therefore the least likely part to develop a detachment. This

paradox has commonly been resolved by postulating that detachments develop on zones of crustal weakness, such as evaporites<sup>19</sup>, shale and marl<sup>14</sup>, or low-viscosity zones related to rheological stratification<sup>20</sup>. Although weak layers could plausibly initiate detachment faults, as demonstrated by the Moho decoupling in Fig. 3b and c, our calculations demonstrate that they are unnecessary. Mid-crustal detachments arise spontaneously from thermo-mechanical feedback effects between the elasto-viscoplastic and brittle layers, even in a homogeneous crustal medium. Furthermore, the results suggest that crustal detachment may be followed by steeply dipping lithosphere-wide shear zones, which ultimately enables lithospheric break-up<sup>21</sup>.

Decoupling of crust and mantle across the Moho is a much discussed issue<sup>3,20,22,23</sup>. However, despite the usage of strong rheological contrasts across the Moho, the development of this decoupling layer has not been recognized in numerical models. Our results illustrate the importance of thermo-mechanical feedback effects for the spontaneous development of this decoupling layer.

The evolution of the lithospheric strength profile, as shown by our model, also resolves a paradox mentioned in ref. 24 with regard to the aseismic behaviour of the upper mantle. According to the 'jelly sandwich' strength model of the continental lithosphere, the upper mantle is expected to be strong and produce earthquakes upon loading. The strong upper mantle should also be found in flexural rigidity analyses of continents. Nonetheless, data show that earthquakes do not occur in the continental upper mantle, and that the latter is commonly not reflected in flexural rigidity analyses. This apparent paradox is resolved in our models, which demonstrate that the mantle evolves from strong to weak as ductile instabilities develop and stabilize in the olivine layer. Although our models do not progress into active seismic events, there is a strong propensity of the quartz layer to develop fast ductile slip reminiscent of slow earthquakes<sup>25</sup>, whereas the olivine layer has the tendency to deform by stable ductile flow.

The energy approach to modelling the evolution of the continental lithosphere shows that its strength profile differs significantly from that derived from the classical quasi-static 'Christmas tree' constructs. This difference hinges on thermo-mechanical feedback effects. Out of a most simple quartz-olivine composite lithospheric composition, the models develop natural features, such as the thickness of the crustal seismogenic zone, mid-crustal detachments at the otherwise strongest crustal layer, lower crustal channel flow above the crust–mantle decoupling zone, and lack of flexural strength or seismicity in the continental upper mantle<sup>10</sup>.

## METHODS

**Model overview.** The approach used here is to solve the fully coupled continuum, momentum and energy equations (Supplementary Appendix 1). This differs from traditional approaches, where the energy equation is not fully coupled or is neglected. We solve the equations for equilibrium and seek the minimum value of free energy<sup>26</sup>, using an implicit adaptive time step technique<sup>26,27</sup>.

**Energy equation.** In order to understand the temporal evolution before and after the initiation of failure, the dissipation of the system must be solved. The master equation that provides the evolution of the dissipation function is the energy equation, with its time derivative. By coupling the energy equation we are able to study the self-consistent history of strain localization through either the brittle or the ductile localization feedback. The strength of our method is that it avoids prescribing shear localization. By doing so, the entropy of the system and the second law of thermodynamics are considered. This contrasts with prescriptive methods, such as the Mohr–Coulomb approach, where, for instance, in order to stabilize shear zones in numerical models, shear weakening after failure is imposed, with no account of the second law of thermodynamics, which it could be violating.

**Modelling brittle and ductile behaviour.** Brittle behaviour in our model differs from classical approaches, which use non-coaxiality in stress and strain rate tensors (non-associative behaviour) by prescribing the dilatancy angle. This implies that localization in itself is prescribed rather than arises spontaneously, and also implies that failure will take place in preferred planes with prescribed

angles. This approach, which was set in the mid-1970s<sup>28,29</sup>, only considers the phenomenon of localization and is not tailored for describing further evolution of deforming systems after failure. For this, energy, and consequently entropy, needs to be considered. In our approach, the brittle regime is solved using standard pressure-dependent yield strength, without referring to the prescriptive Mohr-Coulomb rheology. Instead we use a natural localization phenomenon that results from the full feedback between thermal expansion and (recoverable) temperature changes related to pressure variations (isentropic work; see ref. 30, p. 8).

In our calculations, noise-level strength heterogeneities lead to local thermal expansion due to focusing of work. This leads to a pressure pulse, which further creates thermal expansion and leads to a local departure from coaxiality, bringing the material locally above the yield stress and spontaneous localization of shearing. Our system is thus associative, where pressure controls yielding and thermal expansion controls localization. This approach is particularly faithful to semi-brittle conditions at the brittle-ductile transition, where the dilatancy angle is negligible due to the overburden pressure, and the fracture angle is 45° in plane strain. However, this approach would require specific coupling to the energy evolution of dilatant fractures to appropriately model near surface processes. Hence, our models produce a smoother topography than expected in a natural system.

Ductile deformation also tends to localize into ductile shear zones as a result of feedback between shear heating and thermal softening (Fig. 2). In order to calculate the composite brittle and ductile rheology everywhere, we use the additive strain rate decomposition where temperature- and pressure-dependent strain rates are added. The viscous constitutive law used is power-law. For simplicity, we assume that any strain rates below  $10^{-16} \text{ s}^{-1}$  are elastic.

Received 10 November 2005; accepted 5 May 2006.

- Goetze, C. & Evans, B. Stress and temperature in the bending lithosphere as constrained by experimental rock mechanics. *Geophys. J. R. Astron. Soc.* **59**, 463–478 (1979).
- Brace, F. W. & Kohlstedt, D. L. Limits on lithospheric stress imposed by laboratory experiments. *J. Geophys. Res.* **50**, 6248–6252 (1980).
- Ranalli, G. & Murphy, D. C. Rheological stratification of the lithosphere. *Tectonophysics* **132**, 281–295 (1987).
- Davis, G. H. & Coney, P. J. Geologic development of the Cordilleran metamorphic core complexes. *Geology* **7**, 120–124 (1979).
- Miller, E. L., Gans, P. B. & Garing, J. The Snake Range décollement: an exhumed mid-Tertiary brittle-ductile transition. *Tectonics* **2**, 239–263 (1983).
- Jolivet, L. *et al.* Midcrustal shear zones in postorogenic extension: example from the Tyrrhenian Sea. *J. Geophys. Res.* **103**, 12123–12160 (1998).
- Wernicke, B. in *The Cordilleran Orogen; Conterminous U.S.* (eds Burchfiel, B. C., Lipman, P. W. & Zoback, M. L.) 553–581 (The Geology of North America, Geological Society of America, Boulder, Colorado, 1992).
- Anderson, J. L. in *Metamorphism and Crustal Evolution of the Western United States* (ed. Ernst, W. G.) 502–525 (Prentice Hall, Engelwood Cliffs, New Jersey, 1987).
- Buck, W. R. & Poliakov, A. N. B. Abyssal hills formed by stretching oceanic lithosphere. *Nature* **392**, 272–275 (1998).
- Burov, E. B. & Diament, M. The effective elastic thickness ( $T_e$ ) of continental lithosphere: What does it really mean? *J. Geophys. Res.* **100**, 3905–3927 (1995).
- Ellis, S. & Stöckhert, B. Elevated stresses and creep rates beneath the brittle-ductile transition caused by seismic faulting in the upper crust. *J. Geophys. Res.* **109**, B05407 (2004) doi:10.1029/2003JB002744.
- Shawki, T. G. An energy criterion for the onset of shear localization in thermal viscoplastic material, Part II: applications and implications. *J. Appl. Mech.* **61**, 538–547 (1994).
- Lister, G. S., Banga, G. & Feenstra, A. Metamorphic core complexes of the Cordilleran type in the Cyclades, Aegean Sea. *Geology* **12**, 221–225 (1984).
- Ramsay, G. in *Thrust and Nappe Tectonics* (eds McClay, K. R. & Price, N. J.) 293–309 (Special Publication No. 9, Geological Society, Boulder, Colorado, 1981).
- Cook, F. A. *et al.* LITHOPROBE crustal reflection structure of the southern Canadian Cordillera 1, foreland thrust and fold belt to Fraser River fault. *Tectonics* **11**, 12–35 (1992).
- Wernicke, B. Low-angle normal faults in the Basin and Range province: Nappe tectonics in an extending orogen. *Nature* **291**, 645–648 (1981).
- Gueydan, F., Leroy, Y., Jolivet, L. & Agard, P. Analysis of continental midcrustal strain localization induced by microfracturing and reaction-softening. *J. Geophys. Res.* **108**, 2064 (2003) doi:10.1029/2001JB000611.
- Butler, R. W. H. Thrust tectonics, deep structure and crustal subduction in the Alps and Himalayas. *J. Geol. Soc. Lond.* **143**, 857–873 (1986).
- Laubscher, H. P. *Detachment, Shear and Compression in the Central Alps* 191–211 (Memoir 158, Geological Society of America, Boulder, Colorado, 1983).
- Meissner, R. & Mooney, W. Weakness of the lower continental crust: a condition for delamination, uplift, and escape. *Tectonophysics* **296**, 47–60 (1998).
- Rosenbaum, G., Regenauer-Lieb, K. & Weinberg, R. Continental extension: from core complexes to rigid block faulting. *Geology* **33**, 609–612 (2005).
- Royden, L. H. *et al.* Surface deformation and lower crustal flow in eastern Tibet. *Science* **276**, 788–790 (1997).
- Wobus, C., Heimsath, A., Whipple, K. & Hodges, K. Active out-of-sequence thrust faulting in the central Nepalese Himalaya. *Nature* **434**, 1008–1011 (2005).
- Jackson, J. Strength of the continental lithosphere; time to abandon the jelly sandwich? *GSA Today* **12**, 4–10 (2002).
- Regenauer-Lieb, K. & Yuen, D. Normal-mode excitation by Sumatran earthquake and short-timescale instabilities: Thermo-mechanics and rheology of the Earth's lithosphere. *Eos* **86** (18; Jt. Assem. Suppl.), abstr. U53A–02 (2005).
- Regenauer-Lieb, K. & Yuen, D. A. Positive feedback of interacting ductile faults from coupling of equation of state, rheology and thermal-mechanics. *Phys. Earth Planet. Inter.* **142**, 113–135 (2004).
- ABAQUS/Standard Theory Manual Version 6.1 (Hibbit, Karlsson and Sorenson, Inc., Pawtucket, Rhode Island, 2000).
- Rice, J. R. in *Theoretical and Applied Mechanics* (ed. Koiter, W. T.) 207–220 (North-Holland, Amsterdam, 1977).
- Rudnicki, J. W. & Rice, J. R. Conditions for localization of deformation in pressure-sensitive dilatant materials. *J. Mech. Phys. Solids* **23**, 371–394 (1975).
- Needleman, A. & Tvergaard, V. Analyses of plastic flow deformation in metals. *Appl. Mech. Rev.* **45**, 3–18 (1992).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank V. Lyakhovsky for comments on the manuscript. K.R.-L. acknowledges support from the Johannes Gutenberg-University Mainz and the Predictive Mineral Discovery Cooperative Research Centre.

**Author Information** Reprints and permissions information is available at [npg.nature.com/reprintsandpermissions](http://npg.nature.com/reprintsandpermissions). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to K.R.-L. ([klaus@cyllene.uwa.edu.au](mailto:klaus@cyllene.uwa.edu.au)).

# Earthquakes triggered by silent slip events on Kīlauea volcano, Hawaii

Paul Segall<sup>1</sup>, Emily K. Desmarais<sup>1</sup>, David Shelly<sup>1</sup>, Asta Miklius<sup>2</sup> & Peter Cervelli<sup>3</sup>

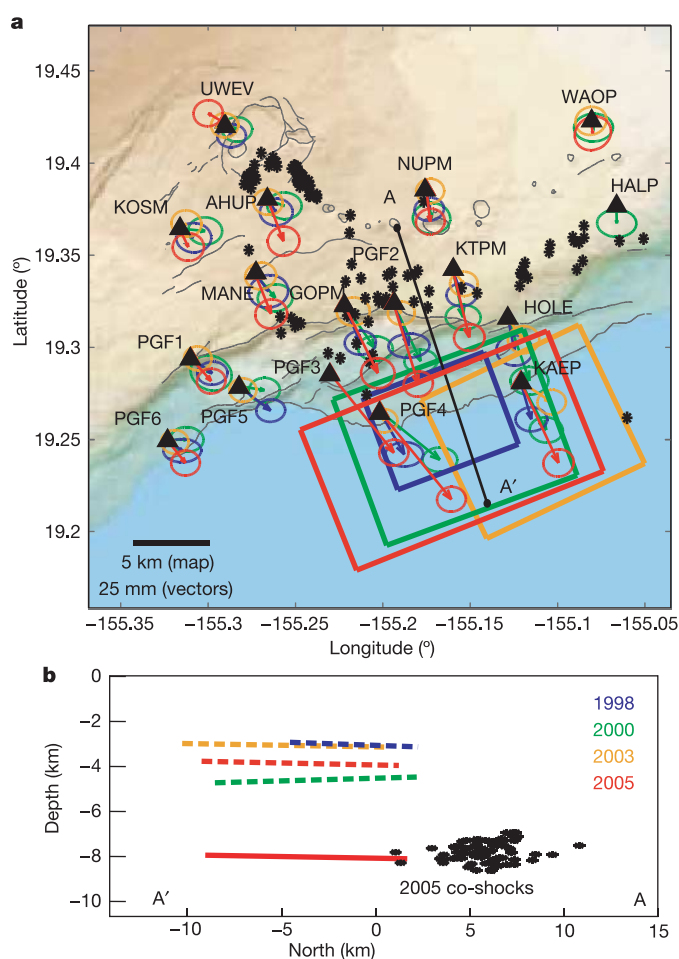
Slow-slip events, or ‘silent earthquakes’, have recently been discovered in a number of subduction zones including the Nankai trough<sup>1–3</sup> in Japan, Cascadia<sup>4,5</sup>, and Guerrero<sup>6</sup> in Mexico, but the depths of these events have been difficult to determine from surface deformation measurements. Although it is assumed that these silent earthquakes are located along the plate megathrust, this has not been proved. Slow slip in some subduction zones is associated with non-volcanic tremor<sup>7,8</sup>, but tremor is difficult to locate and may be distributed over a broad depth range<sup>9</sup>. Except for some events on the San Andreas fault<sup>10</sup>, slow-slip events have not yet been associated with high-frequency earthquakes, which are easily located. Here we report on swarms of high-frequency earthquakes that accompany otherwise silent slips on Kīlauea volcano, Hawaii. For the most energetic event, in January 2005, the slow slip began before the increase in seismicity. The temporal evolution of earthquakes is well explained by increased stressing caused by slow slip, implying that the earthquakes are triggered. The earthquakes, located at depths of 7–8 km, constrain the slow slip to be at comparable depths, because they must fall in zones of positive Coulomb stress change. Triggered earthquakes accompanying slow-slip events elsewhere might go undetected if background seismicity rates are low. Detection of such events would help constrain the depth of slow slip, and could lead to a method for quantifying the increased hazard during slow-slip events, because triggered events have the potential to grow into destructive earthquakes.

A silent earthquake beneath the south flank of Kīlauea volcano on 10–11 November 2000 displaced Global Positioning System (GPS) stations as much as 1.5 cm over about 36 hours<sup>11</sup>. The depth of the subhorizontal fault was not well constrained, but inversions favoured depths of 4–5 km, considerably shallower than the decollement thought to occur at the base of the volcano. We now recognize similar events on 20–21 September 1998, 3–4 July 2003, and 26–27 January 2005 (ref. 12 and additional events have been reported<sup>13</sup>). All four events have similar durations and displacement patterns (Figs 1 and 2). Inversions assuming uniform slip dislocations place the four sources in virtually the same location (Fig. 1). Whereas the November 2000 slow slip was preceded by extreme rainfall<sup>11</sup>, the other events were not.

All four slow-slip events were associated with heightened levels of microseismicity (Fig. 2). The cumulative magnitude of the micro-earthquakes is far too small to explain the observed displacements. For example, the cumulative moment of the 2005 earthquake swarm is  $\sim 1.8 \times 10^{14}$  N m, far less than that of the slow slip,  $6.8 \times 10^{17}$  N m. The microearthquakes, concentrated adjacent to the landward edge of the dislocation (Fig. 1), are thus not the source of the deformation.

The association of high-frequency earthquakes with slow slip could be explained by either (1) the earthquakes unpinning the fault, allowing slow slip to occur, or (2) the slow slip stressing the adjacent fault, thereby increasing the seismicity rate. To constrain

the onset and duration of fault slip relative to the microearthquakes, we invert the GPS observables during the 2005 slow event directly for fault slip as a function of time<sup>11</sup>. The slow slip started early on 26 January 2005, well before the dramatic increase in seismicity, and

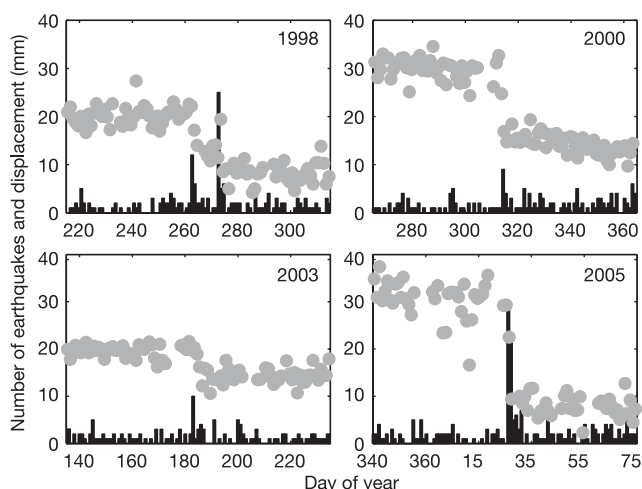


**Figure 1 | Displacements and inferred slip zones for four silent slip events.** **a**, Map view. Vectors indicate displacements determined as the difference between the mean position before and after the event. Ellipses represent 95% confidence intervals. Rectangles show surface projections of best-fitting dislocations found by non-linear optimization, using a simulated annealing procedure<sup>27</sup>. Asterisks indicate relocated earthquakes accompanying the 2005 slip event. **b**, Cross-section. Dashed lines represent dislocations from inversion of GPS-derived displacements. The solid red line indicates the 2005 event with depth constrained by seismicity (see text). GPS station locations (triangles) indicated with abbreviations.

<sup>1</sup>Geophysics Department, Stanford University, Stanford, California 94305, USA. <sup>2</sup>USGS Hawaii Volcano Observatory, PO Box 51 Hawaii National Park, Hawaii 96718-0051, USA.

<sup>3</sup>USGS Alaska Volcano Observatory, 4200 University Drive, Anchorage, Alaska 99508, USA.





**Figure 2 | Temporal association of deformation and seismicity.** North component of displacement of GPS station KAEP (grey circles) and number of Kilauea south-flank earthquakes per day (black histogram). Note that the seismicity rate increases during periods of rapid displacement.

continued for approximately two days (Fig. 3), supporting the second interpretation. The triggered earthquakes should thus properly be thought of as ‘co-shocks’ and aftershocks of the otherwise silent earthquakes.

Dieterich’s seismicity rate theory<sup>14</sup> is used to quantitatively relate the slip and seismicity. The seismicity rate  $R$  is related to the background seismicity rate  $r$  and a state variable  $\gamma$ , as:

$$R = \frac{dN}{dt} = \frac{r}{\gamma \dot{\tau}_r} \quad (1)$$

where  $N$  is the number of events, and  $\dot{\tau}_r$  is the background stressing rate. The seismicity state variable evolves according to:

$$d\gamma = \frac{1}{a\sigma} [dt - \gamma d\tau + \gamma(\tau/\sigma - \alpha)d\sigma] \quad (2)$$

where  $a$  and  $\alpha$  are constitutive constants, and  $\tau$  and  $\sigma$  are the shear and effective normal stresses, respectively. Because normal stress variations may be largely balanced by undrained changes in pore pressure, we assume  $d\sigma = 0$ ; in fact the stress variations are understood to be changes in the Coulomb stress. Dieterich<sup>14</sup> showed that the seismicity rate following a step change in shear stress  $\Delta\tau$ , followed by a return to the background stressing rate  $\dot{\tau}_r$ , yields the modified Omori law with aftershock duration given by  $t_a = a\sigma/\dot{\tau}_r$ .

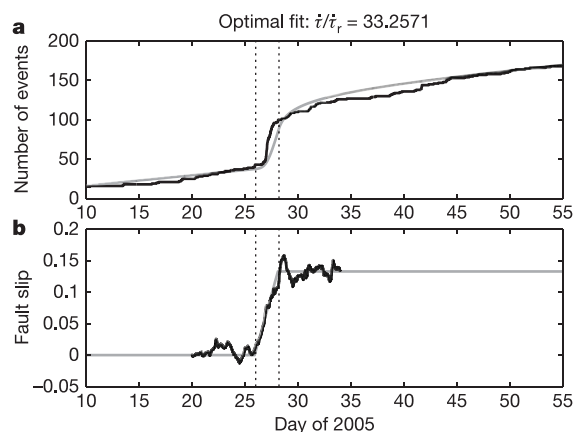
To model the triggered seismicity, we approximate the slip history with a ramp function (Fig. 3b). Before the onset of accelerated slip,  $t < t_0$ , the background stressing rate is  $\dot{\tau}_r$ . During the event  $t_0 < t < t_1$  the stressing rate increases to  $\dot{\tau}$ ; for  $t > t_1$  the stressing rate returns to background. For this stress history the predicted seismicity rate is found from equations (1) and (2) (see Methods). To compare with observations we compute the cumulative number of earthquakes,  $N(t)$ , determined by integrating  $R(t)$ :

$$N(t) = \begin{cases} rt_a \ln \left[ \frac{\dot{\tau}}{\dot{\tau}_r} \left( \exp \left( \frac{\dot{\tau}(t-t_0)}{a\sigma} \right) - 1 \right) + 1 \right] & \text{for } t_0 < t < t_1 \\ rt_a \ln \left[ \frac{\exp \left( \frac{\dot{\tau}(t-t_1)}{a\sigma} \right) + C}{1+C} \right] & \text{for } t > t_1 \end{cases} \quad (3)$$

where

$$C = \left[ \left( 1 - \frac{\dot{\tau}_r}{\dot{\tau}} \right) \exp \left( -\frac{\dot{\tau}(t_1 - t_0)}{a\sigma} \right) + \frac{\dot{\tau}_r}{\dot{\tau}} - 1 \right] \quad (4)$$

$N(t)$  depends on five parameters: (1) the background rate  $r$ , (2) the aftershock decay time,  $t_a$ ; (3) the ratio of the event to the background stressing rate,  $\dot{\tau}/\dot{\tau}_r$ , (4) the onset  $t_0$  and (5) the duration  $t_1 - t_0$  of the slip event. Note that  $\dot{\tau}/a\sigma = \dot{\tau}/\dot{\tau}_r t_a$ .



**Figure 3 | Cumulative number of earthquakes compared to that predicted for a slip event by equation (3).** **a**, Observed earthquake count in hourly bins (black line); predicted for the slow-slip event (grey line). Vertical dotted lines mark the beginning and end of the slow-slip event. **b**, Inverted slip history estimated directly from the GPS phase data (black line), using a Kalman filter procedure<sup>11</sup>. Fault slip is allowed to vary as a random walk in time, with scale parameter  $\sigma_s = 0.015 \text{ mm yr}^{-1/2}$ . The ramp function illustrates the stress history used to derive the predicted seismicity rate (grey line).

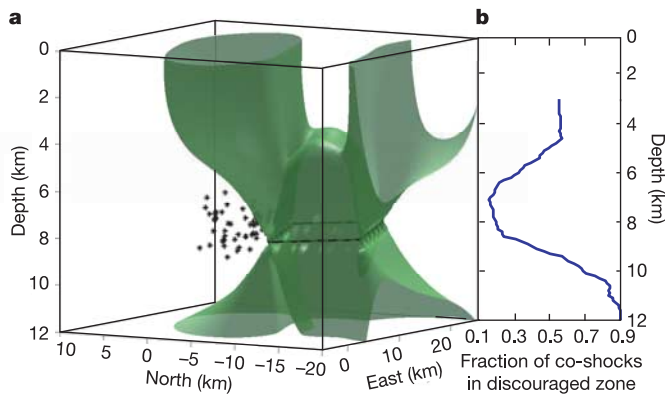
The background rate  $r$  of  $\sim 1.33$  events per day is estimated from the Hawaii Volcano Observatory (HVO) catalogue before and well after the slow-slip event. The onset time (GMT midnight on 26 January) and duration (2.2 days) of the slow event are determined from the GPS data (Fig. 3b). A  $t_a$  of 10 to 20 days is found from the decay of aftershocks following the 30 June 1997, magnitude  $M_w = 5.5$  south-flank earthquake. The only parameter not determined *a priori* is the ratio of stressing rates  $\dot{\tau}/\dot{\tau}_r$ .

We fitted the cumulative number of earthquakes to equation (3) with  $t_a = 10$ . The best fit, obtained for an increase in stressing rate of a factor of 33, provides satisfactory agreement with the earthquake data (Fig. 3), especially considering the single adjustable parameter. A better fit is obtained by reducing  $t_a$  to seven days. Whether this indicates temporal or spatial variation in  $t_a$  (the 1997 earthquake was roughly 20 km from the swarm earthquakes) is unknown.

The spatial association of the silent slip and its co-shocks is clear when viewed on a map, but the depth of the slow-slip event is difficult to constrain solely on the basis of the GPS observations. Although the catalogue earthquake depths are scattered over a broad range, relocations of south-flank earthquakes illuminate a sub-horizontal plane<sup>15,16</sup>. Hansen *et al.*<sup>17</sup> used a temporary deployment of 29 three-component seismometers along with the HVO permanent network jointly to locate earthquakes and determine the three-dimensional seismic velocity structure. They found that earthquakes occurring on the central south flank from November 1999 to June 2000 lie on a nearly horizontal surface at a depth of 7 to 9 km.

The Hansen *et al.*<sup>17</sup> hypocentres can be used to improve the locations of the swarm events accompanying slow slips. We focus on the most energetic January 2005 swarm. From their catalogue locations, we infer that quakes during the other slow events are located at comparable depths. A ‘double difference’ relocation<sup>18</sup> of the January 2005 swarm events relative to the 1999–2000 events (see Methods) demonstrates that the swarm events were located at the same depth as the background seismicity, 7 to 9 km (Fig. 1).

The depth of the swarm earthquakes, and the likelihood that they were triggered by the slow slip, constrains the depth of the slow slip. Specifically, slow slip must have occurred at depths for which the induced stresses favour slip in the swarm. If the slow slip is too shallow the earthquakes locate in a stress shadow and are thus inconsistent with triggering. Varying the depth of the best-fitting dislocation maps the depth range consistent with the triggered



**Figure 4 | Earthquake locations in relation to Coulomb stress change due to fault slip.** **a**, January 2005 earthquakes (asterisks) and isosurface of constant Coulomb stress change (CSC) on horizontal planes, with dislocation at 8 km depth as indicated by dashed slice. Outside the green surface the stress change encourages slip. For a dislocation surface at 4–5 km depth, a majority of the earthquakes fall within the zone of negative CSC. **b**, Fraction of earthquakes for which  $CSC < 0$  as a function of the depth of the slow-slip event. The minimum at a depth of 6.5 to 8.5 km indicates the preferred depth of the slow-slip zone.

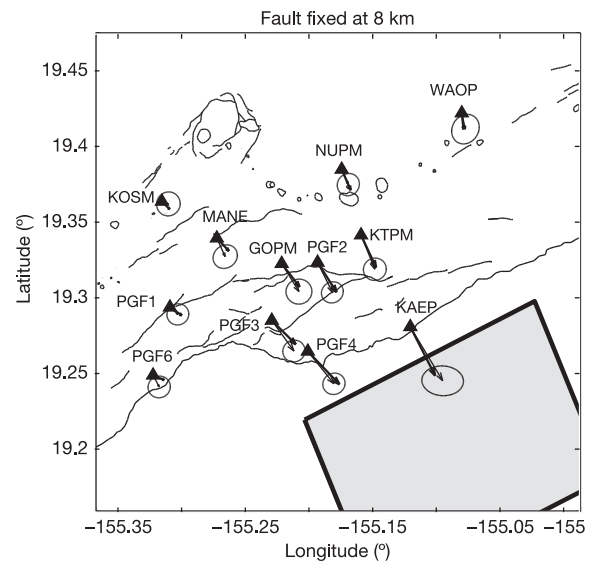
earthquakes (Fig. 4). The best fit occurs when the slip event is at the same depth as the earthquakes, 6.5–8.5 km. At this depth the shear stress concentration at the edge of the dislocation is focused on the earthquake swarm. The silent slip event and its co-shocks are thus probably coplanar and located at a depth of 6.5–8.5 km.

Is the depth of the silent slip event inferred from the earthquakes consistent with the geodetic observations? The best-fitting dislocation constrained to a depth of 8.0 km (Fig. 5) is in fact consistent with the data at the few-millimetre level. Depth-varying elastic properties and non-planar topography favour deeper sources relative to uniform half-space models, although the effect is relatively minor.

Kilauea suffered an  $M_w = 7.7$  earthquake and tsunami in 1975 (ref. 19). Despite the fact that both geodetic data<sup>20</sup> and the tsunami source<sup>21</sup> require slip offshore, the aftershocks were restricted to a narrow strip between the rift zones and the coastline. Indeed, south-flank earthquakes rarely occur offshore, despite the fact that the geodetic data require extensive slip there<sup>22,23</sup>. These observations indicate a transition from stick-slip behaviour between the rift zone and the coast to stable sliding offshore. Slow-slip events seem to occur in the transition between these two domains. Modelling studies indicate that transient slip events occur in transitions from velocity weakening to velocity strengthening friction<sup>24</sup>, or where velocity-weakening patches in an otherwise creeping fault are near the critical nucleation dimension<sup>25</sup>. On Kilauea the transition in frictional behaviour might result from temperature and pressure variations with distance from the rift zone<sup>26</sup>.

Our results have implications for silent slip events in subduction zones. (A slow earthquake on the San Andreas fault<sup>10</sup> appears to have been triggered in part by a sequence of  $M_w = 3+$  earthquakes, which initiated two hours before detectable strain changes, and is thus very different from the Kilauea slow events.) Microseismicity rates on Kilauea are much higher than in some subduction zones. A rate increase of a factor of 35 is dramatic on Kilauea, but might go unnoticed in subduction zones with few earthquakes located on the plate interface. Given our findings, a concerted effort should be made to search for very small earthquakes accompanying slow-slip events elsewhere.

Slow-slip events in subduction zones appear to be located down-dip of the locked zone, so that transient slip acts to stress the seismogenic fault. If small events are triggered, as we observe on Kilauea, then the potential exists for one of these to grow into a destructive earthquake. A  $M_w = 6.7$  thrust earthquake coincided



**Figure 5 | Dislocation at a depth of 8 km fits the GPS data well.** Observed (thin vectors with 95% confidence ellipses) and predicted (bold vectors) displacements during the January 2005 silent slip event with dislocation constrained to depth of 8.0 km. Surface projection of best fitting dislocation shown as rectangle.

with the end of the 2002 slow-slip event in Guerrero, Mexico<sup>24</sup>. It is possible to quantify the increased hazard associated with slow-slip events by the increase in seismicity rate, which depends on the duration of the slow slip relative to  $t_a$  and the increase in stressing rate  $\dot{\tau}/\dot{\tau}_r$ . The peak seismicity rate  $1/(C+1)$  occurs at the end of the slow-slip event  $t_1$  (see Methods). Of course, the nucleation of an earthquake does not determine its ultimate size. The longer the plate boundary remains locked, however, the higher the ambient stresses become, and the more likely it is that a triggered event will grow into a major earthquake. It is possible that as the stress increases the size of co-shocks triggered during slow events will increase, making them more easily detected.

## METHODS

**Predicted seismicity rate.** At  $t = 0$ ,  $\gamma$  takes the value  $1/\dot{\tau}_r$ . Ignoring changes in effective normal stress, for  $0 < t < t_1$  the stressing rate is constant at  $\dot{\tau}$ , so that equation (2) reduces to:

$$\frac{d\gamma}{dt} = \frac{1}{a\sigma} [1 - \gamma\dot{\tau}] \quad (5)$$

which has solution:

$$\gamma = \left( \frac{1}{\dot{\tau}_r} - \frac{1}{\dot{\tau}} \right) \exp \left( -\frac{\dot{\tau}t}{a\sigma} \right) + \frac{1}{\dot{\tau}} \quad \text{for } 0 < t < t_1 \quad (6)$$

given the initial condition  $\gamma(t=0) = 1/\dot{\tau}_r$ . For  $t > t_1$  the stressing rate is again constant, but at the background rate. The solution to equation (5) is thus of a similar form to equation (6), but the initial condition is now given by  $\gamma_1 \equiv \gamma(t_1)$ , that is, equation (6) evaluated at  $t_1$ . The result is:

$$\gamma = \left( \gamma_1 - \frac{1}{\dot{\tau}_r} \right) \exp \left( -\frac{\dot{\tau}_r(t-t_1)}{a\sigma} \right) + \frac{1}{\dot{\tau}_r} \quad \text{for } t > t_1 \quad (7)$$

The seismicity rate can now be calculated simply from equation (1).

$$\frac{R(t)}{r} = \begin{cases} \left[ \left( 1 - \frac{\dot{\tau}_r}{\dot{\tau}} \right) \exp \left( -\frac{\dot{\tau}(t-t_0)}{a\sigma} \right) + \frac{\dot{\tau}_r}{\dot{\tau}} \right]^{-1} & t_0 < t < t_1 \\ \left\{ C \exp \left( -\frac{\dot{\tau}(t-t_1)}{a\sigma} \right) + 1 \right\}^{-1} & t > t_1 \end{cases} \quad (8)$$

We note that if the duration of the slip event is long compared to  $a\sigma/\dot{\tau}$  the seismicity rate approaches a steady state that is a factor of  $\dot{\tau}/\dot{\tau}_r$  over the background rate. Following the event, as  $t \rightarrow \infty$ , the seismicity rate returns to background. We also note that as  $\dot{\tau} \rightarrow \infty$  and  $(t_1 - t_0) \rightarrow 0$  such that the product  $\dot{\tau}(t_1 - t_0) \rightarrow \Delta\tau$ , equation (8) reduces to equation (12) in ref. 14, which gives the Omori-like decay of events following a step change in stress.

To estimate the aftershock duration  $t_a$  we fitted the cumulative number of earthquakes following a high-frequency mainshock. Dieterich<sup>14</sup> found that the number of earthquakes following a step change in shear stress  $\Delta\tau$  is given by:

$$N(t) = rt_a \ln\{e^{\Delta\tau/a\sigma}(e^{t/t_a} - 1) + 1\} \quad (9)$$

**Relative earthquake locations.** The data are arrival times for the 2005 events, determined by the Hawaiian Volcano Observatory, and arrival times for the 1999–2000 events, provided by the University of Wisconsin. We used a one-dimensional layered velocity model that approximates the three-dimensional model of ref. 17. The 1999–2000 events were fixed at the locations determined by Hansen *et al.*<sup>17</sup>. To minimize potential bias from the one-dimensional model, we weighted the differential times between the 1999/2000 and 2005 events three times more heavily than the remaining differential times.

Received 11 January; accepted 23 May 2006.

- Hirose, H., Hirahara, K., Kimata, F., Fujii, N. & Miyazaki, S. A slow thrust slip event following the two 1996 Hyuganada earthquakes beneath the Bungo Channel, southwest Japan. *Geophys. Res. Lett.* **26**, 3237–3240 (1999).
- Ozawa, S. *et al.* Detection and monitoring of ongoing aseismic slip in the Tokai region, central Japan. *Science* **298**, 1009–1012 (2002).
- Miyazaki, S., McGuire, J. & Segall, P. A transient subduction zone slip episode in southwest Japan observed by the nationwide GPS array. *J. Geophys. Res.* **108**, 2087–2087 (2003).
- Dragert, H., Wang, K. L. & James, T. S. A silent slip event on the deeper Cascadia subduction interface. *Science* **292**, 1525–1528 (2001).
- Miller, M. M., Melbourne, T., Johnson, D. J. & Sumner, W. Q. Periodic slow earthquakes from the Cascadia subduction zone. *Science* **295**, 2423–2423 (2002).
- Kostoglodov, V. *et al.* A large silent earthquake in the Guerrero seismic gap, Mexico. *Geophys. Res. Lett.* **30**, 1807–1810 (2003).
- Rogers, G. & Dragert, H. Episodic tremor and slip on the Cascadia subduction zone: The chatter of silent slip. *Science* **300**, 1942–1943 (2003).
- Obara, K., Hirose, H., Yamamizu, F. & Kasahara, K. Episodic slow slip events accompanied by non-volcanic tremors in southwest Japan subduction zone. *Geophys. Res. Lett.* **31**, L23602, doi:10.1029/2004GL020848 (2004).
- Kao, H. *et al.* A wide depth distribution of seismic tremors along the northern Cascadia margin. *Nature* **436**, 841–844 (2005).
- Linde, A., Gladwin, M., Johnston, M., Gwyther, R. & Bilham, R. A slow earthquake sequence on the San Andreas fault. *Nature* **383**, 65–68 (1996).
- Cervelli, P., Segall, P., Johnson, K., Lisowski, M. & Miklius, A. Sudden aseismic fault slip on the south flank of Kilauea volcano. *Nature* **415**, 1014–1018 (2002).
- Segall, P., Desmarais, E., Miklius, A. & Okubo, P. (Nearly) silent earthquakes on volcanoes. *Trans. AGU Fall Meet. Suppl.* **86**(52), G43A-01 (2005).
- Brooks, B., Foster, J., Bevis, M., Frazer, L. & Behn, M. Slow earthquakes on the flank of Kilauea volcano, Hawaii. *Trans. AGU Fall Meet. Suppl.* **86**(52), G53B-0879 (2005).
- Dieterich, J. A constitutive law for rate of earthquake production and its application to earthquake clustering. *J. Geophys. Res.* **99**, 2601–2618 (1994).
- Got, J., Frechet, J. & Klein, F. Deep fault plane geometry inferred from multiplet relative relocation beneath the south flank of Kilauea. *J. Geophys. Res.* **99**, 15375–15386 (1994).
- Got, J.-L. & Okubo, P. New insights into Kilauea's volcano dynamics brought by large scale relative relocation of microearthquakes. *J. Geophys. Res.* **108**, 2337–2350 (2003).
- Hansen, S., Thurber, C., Mandernach, M., Haslinger, F. & Doran, C. Seismic velocity and attenuation structure of the East Rift Zone and south flank of Kilauea volcano, Hawaii. *Bull. Seismol. Soc. Am.* **94**, 1430–1440 (2004).
- Waldhauser, F. & Ellsworth, W. L. A double-difference earthquake location algorithm: method and application to the Northern Hayward Fault, California. *Bull. Seismol. Soc. Am.* **90**, 1353–1368 (2000).
- Nettles, M. & Ekstrom, G. Long-period source characteristics of the 1975 Kalapana, Hawaii, earthquake. *Bull. Seismol. Soc. Am.* **94**, 422–429 (2004).
- Owen, S. & Burgmann, R. An increment of volcano collapse: Kinematics of the 1975 Kalapana, Hawaii, earthquake. *J. Volcan. Geotherm. Res.* **150**, 163–185 (2006).
- Ma, K.-F., Kanamori, H. & Satake, K. Mechanism of the 1975 Kalapana, Hawaii, earthquake inferred from tsunami data. *J. Geophys. Res.* **104**, 13153–13168 (1999).
- Delaney, P. *et al.* Volcanic spreading at Kilauea, 1976–1996. *J. Geophys. Res.* **103**, 18003–18023 (1998).
- Owen, S. *et al.* Rapid deformation of Kilauea volcano: Global positioning system measurements between 1990 and 1996. *J. Geophys. Res.* **105**, 18983–18998 (2000).
- Liu, Y. & Rice, J. R. Aseismic slip transients emerge spontaneously in 3d rate and state modeling of subduction earthquake sequences. *J. Geophys. Res.* **110**, B08307, doi:10.1029/2004JB003424 (2005).
- Kato, N. Interaction of slip on asperities: Numerical simulation of seismic cycles on a two-dimensional planar fault with nonuniform frictional property. *J. Geophys. Res.* **109**, B12306, doi:10.1029/2004JB003001 (2004).
- Segall, P., Cervelli, P. & Miklius, A. Insights from deformation during the Pu'u O'o eruption of Kilauea volcano. *Eos Trans. AGU Fall Meet. Suppl.* **83**(47), abstr. V62C-02 (2002).
- Cervelli, P., Murray, M., Segall, P., Aoki, Y. & Kato, T. Estimating source parameters from deformation data, with an application to the March 1997 earthquake swarm off the Izu Peninsula, Japan. *J. Geophys. Res.* **106**, 11217–11238 (2001).

**Acknowledgements** We gratefully acknowledge C. Thurber for providing seismic data from the University of Wisconsin 1999/2000 experiment, P. Okubo for help with HVO seismic data, the University of Hawaii for exchange of GPS data, and D. Swanson for editorial comments.

**Author Contributions** P.S. and E.K.D. jointly modelled GPS and seismic data. P.S. derived the seismicity rate results. D.S. relocated the earthquakes. A.M. processed GPS data and first detected the events studied. P.C. helped with time-dependent slip inversion. All authors discussed results and commented on manuscript.

**Author Information** Reprints and permissions information is available at [npg.nature.com/reprintsandpermissions](http://npg.nature.com/reprintsandpermissions). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.S. ([segall@stanford.edu](mailto:segall@stanford.edu)).



# Local migration promotes competitive restraint in a host-pathogen 'tragedy of the commons'

Benjamin Kerr<sup>1</sup>, Claudia Neuhauser<sup>2</sup>, Brendan J. M. Bohannan<sup>3</sup> & Antony M. Dean<sup>2</sup>

Fragmented populations possess an intriguing duplicity: even if subpopulations are reliably extinction-prone, asynchrony in local extinctions and recolonizations makes global persistence possible<sup>1–8</sup>. Migration is a double-edged sword in such cases: too little migration prevents recolonization of extinct patches, whereas too much synchronizes subpopulations, raising the likelihood of global extinction. Both edges of this proverbial sword have been explored by manipulating the rate of migration within experimental populations<sup>1,3–6,8</sup>. However, few experiments have examined how the evolutionary ecology of fragmented populations depends on the pattern of migration<sup>5</sup>. Here, we show that the migration pattern affects both coexistence and evolution within a community of bacterial hosts (*Escherichia coli*) and viral pathogens (T4 coliphage) distributed across a large network of subpopulations. In particular, different patterns of migration select for distinct pathogen strategies, which we term 'rapacious' and 'prudent'. These strategies define a 'tragedy of the commons': rapacious phage displace prudent variants for shared host resources, but prudent phage are more productive when alone. We find that prudent phage dominate when migration is spatially restricted, while rapacious phage evolve under unrestricted migration. Thus, migration pattern alone can determine whether a *de novo* tragedy of the commons is resolved in favour of restraint.

Biological communities are often fragmented into subpopulations that are loosely linked through migration<sup>10</sup>. The dynamics of such 'metapopulations' have been investigated using theoretical models, statistical tools, field studies and laboratory experiments<sup>10–13</sup>. While extremely valuable, experiments with multi-species metapopulations have focused on ecological dynamics (for example, abundance data) in relatively small networks. The experimental evolution of interacting species in large metapopulations remains comparatively unexplored.

Using a model host–pathogen system<sup>14–17</sup>, we investigated how the pattern of migration within large metapopulations affects eco-evolutionary dynamics. The bacterial host (*E. coli*) and its viral pathogen (T4) were embedded in 96-well microtitre plates, which imposed a metapopulation structure: each media-filled well was a subpopulation nested within the larger collection. The entire metapopulation was perpetuated through serial transfer, with each well culture diluted into a well with fresh medium every 12 h. At each transfer, migrations between wells were executed by a high-throughput liquid-handling robot according to a pre-specified migration scheme.

Our system illustrates the duplicitous nature of metapopulations. Empirically we find that bacteria and phage cannot coexist within a single well. However, our empirically calibrated simulations predict that coexistence in the metapopulation is feasible (see Box 1 and Fig. 1a, b). Wells come in three varieties (bacteria-filled, phage-filled and empty/media-filled) and exhibit a kind of 'rock–paper–scissors' dynamic: Migrating phage 'beat' bacteria (through infection and decimation), migrating bacteria 'beat' empty wells (through colonization),

and empty wells 'overtake' phage (without host, phage are diluted to extinction). Spatially restricted migration leads to clumping of well types (Fig. 1c), which indicates spatial synchrony at small scales. The spatial asynchrony at large scales, which leads to the dampening of inherent oscillatory dynamics (compare Fig. 1a and Fig. 1b), is due to the decay of autocorrelations over space inherent to restricted

## Box 1 | Empirically calibrated stochastic cellular automata

Each well in the metapopulation can be characterized as being in one of five discrete states: (1) filled with bacteria (state B), (2)–(4) filled with phage at one of three titres (decreasing from states P1 to P2 to P3) and no bacteria or (5) empty (state E). Even with abundant host, phage below the P3 titre have insufficient reproduction to compensate for dilution, and are thus effectively extinct. We empirically determine how each of these well states is transformed under dilution and migration events (see Methods). Collecting this information on single well transformations into a transition matrix (see Box 1 Table 1), we can explore the dynamics of a large collection of wells using stochastic cellular automata.

In this simulation technique, each point in a lattice represents a single well in one of the five possible states. The states of all points at time  $t$  are represented by  $S_t$ . Randomly assigning one of the states B, P1 or E to each point produces  $S_0$ . For each time step, every point in  $S_t$  is updated synchronously according to the rules in Table 1 to produce  $S_{t+1}$ . Migration into any focal point occurs with probability  $m$  from a randomly chosen point in its 'neighbourhood'. The 'Restricted neighbourhood' is the set of points directly north, south, east and west of the focal point (we employ wrap-around boundaries to avoid edge effects). The 'Unrestricted neighbourhood' is the set of all points in the lattice, minus the focal point. In Supplementary Information III we relate our stochastic cellular automata to another approach, the configuration field approximation.

## Box 1 Table 1 | The transition matrix for states of wells

| Current state | Future state      |              |                            |                        |                    |
|---------------|-------------------|--------------|----------------------------|------------------------|--------------------|
|               | Bacteria (B)      | Phage 1 (P1) | Phage 2 (P2)               | Phage 3 (P3)           | Empty (E)          |
| Bacteria (B)  | B, E, $\emptyset$ | P1, P2, P3   | NA                         | NA                     | NA                 |
| Phage 1 (P1)  | NA                | B            | P1, P2, P3, E, $\emptyset$ | NA                     | NA                 |
| Phage 2 (P2)  | NA                | B            | P1                         | P2, P3, E, $\emptyset$ | NA                 |
| Phage 3 (P3)  | NA                | B            | P1                         | P2                     | P3, E, $\emptyset$ |
| Empty (E)     | B                 | NA           | P1                         | P2                     | P3, E, $\emptyset$ |

All entries in this table give the identity of the state of the source well of migration that allows a focal well to transition from the relevant row state to the relevant column state. A well in state B has approximately  $3.0 \times 10^7$  bacterial cells  $\text{ml}^{-1}$ , a well in state P1 has more than  $3.0 \times 10^7$  phage  $\text{ml}^{-1}$ , a well in state P2 has between  $3.0 \times 10^6$  and  $3.0 \times 10^7$  phage  $\text{ml}^{-1}$ , a well in state P3 has between  $3.0 \times 10^5$  and  $3.0 \times 10^6$  phage  $\text{ml}^{-1}$ , and state E corresponds to a well with no bacteria and fewer than  $3.0 \times 10^5$  phage  $\text{ml}^{-1}$ . The symbol  $\emptyset$  refers to transitions in which there is no migration event (just a straight dilution of a focal well followed by incubation). NA, transitions that cannot occur.

<sup>1</sup>Department of Biology, University of Washington, Box 351800, Seattle, Washington 98195-1800, USA. <sup>2</sup>Department of Ecology, Evolution and Behaviour, University of Minnesota, 100 Ecology, 1987 Buford Circle, St Paul, Minnesota 55108, USA. <sup>3</sup>Department of Biological Sciences, Stanford University, Stanford, California 94305, USA.

migration. With spatially restricted migration, only phage at the boundaries between phage clumps and bacterial clumps can access fresh host. Relative to unrestricted migration, this limitation on host access is predicted to reduce the average phage density and increase the average bacterial density (compare Fig. 1a and Fig. 1b).

We tested these theoretical predictions using experimental metapopulations of 192 subpopulations (two 96-well plates) propagated by serial transfer on a 12-h cycle. Immediately after transfer, each metapopulation was exposed to one of three migration treatments. In the 'Restricted' treatment, migration into a focal well occurred from one of the north, south, east or west neighbours with probability  $m = 0.45$  (we used 'wrap-around' boundaries so that every well had four neighbours). In the 'Unrestricted' treatment, migration into a focal well occurred from one of any other well in the metapopulation with the same probability,  $m = 0.45$  (see Supplementary Information IV for the rationale behind this choice of migration probability, as well as more specific theoretical predictions). Although they varied with regard to the pattern of migration, both Restricted and Unrestricted treatments imposed some degree of population structure. In the 'Well-Mixed' treatment, all population structure was destroyed every 12-h cycle by thoroughly mixing the entire metapopulation in a large reservoir and then redistributing the mixture into a fresh set of wells. Each treatment was run for 20 cycles and replicated four times.

Consistent with theoretical predictions, we observed that population structure is critical to coexistence. The Well-Mixed treatment lost both bacteria and phage, whereas both types persisted in the other treatments (see Fig. 2a, b). Population instability was gauged by the coefficient of variation (CV) in phage or bacterial abundance over all time points. By this measure, bacteria in the Restricted treatment were more stable than in the Unrestricted treatment, although not significantly so ( $\overline{CV}_{\text{Restricted}} = 0.951$ ,  $\overline{CV}_{\text{Unrestricted}} = 1.276$ ,  $P = 0.2$ ). Phage in the Restricted treatment were significantly more stable than in the Unrestricted treatment ( $\overline{CV}_{\text{Restricted}} = 0.572$ ,  $\overline{CV}_{\text{Unrestricted}} = 0.884$ ,  $P = 0.02857$ ). Thus, restricted migration tends to stabilize community dynamics, as predicted by our simulations. In contrast

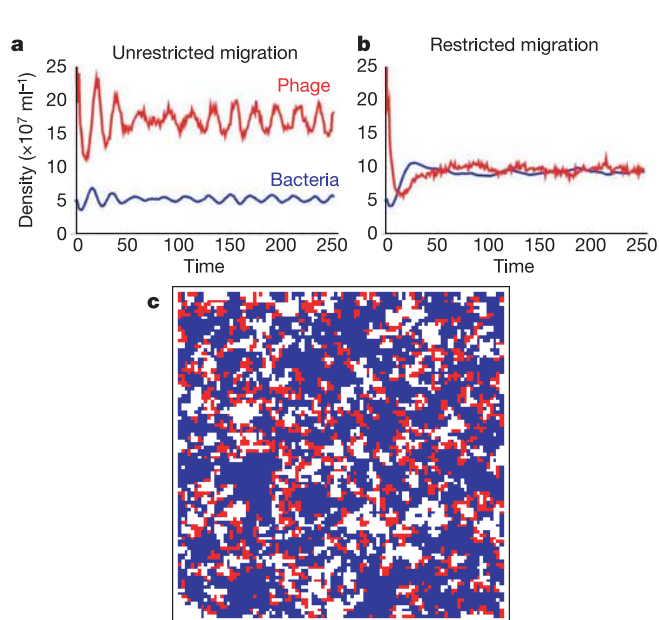
to our predictions, we found no significant differences across treatments in average bacterial density ( $\overline{N}_{\text{Restricted}} = 4.18 \times 10^7$ ,  $\overline{N}_{\text{Unrestricted}} = 4.03 \times 10^7$ ,  $P = 0.89$ ) or in average phage density ( $\overline{N}_{\text{Restricted}} = 2.5 \times 10^8$ ,  $\overline{N}_{\text{Unrestricted}} = 2.76 \times 10^8$ ,  $P = 0.34$ ). What might account for this discrepancy?

The early appearance of resistant bacteria in the Unrestricted treatment might lower phage density below predicted values. However, appreciable resistance appeared only at the end of the experiment and did not differ in frequency between the treatments. As further evidence against the role of host evolution, bacterial isolates from the last transfer of each treatment did not differ in fitness from their common ancestor (see Supplementary Information I). By contrast, a stable polymorphism was evident in the phage after only a few transfers (see Supplementary Information II). Could evolution in the phage (rather than its bacterial host) lead to a relatively lower average phage density in the Unrestricted treatment?

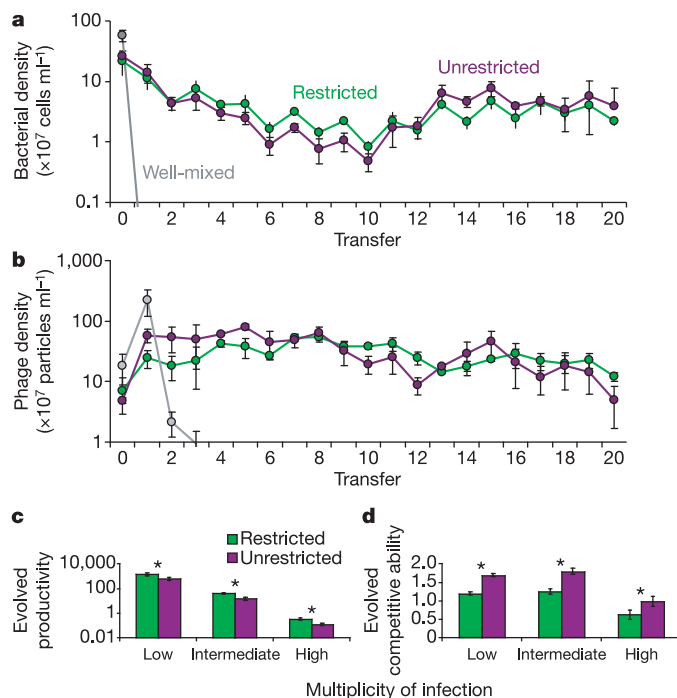
To explore this possibility, we sampled four random phage isolates from the last transfer of each replicate and determined their productivities (number of phage produced per parent). Regardless of the initial multiplicity of infection (MOI, or ratio of phage to bacteria), phage from the Unrestricted treatment were significantly less productive than phage from the Restricted treatment (Fig. 2c). Hence, evolution in the phage accounts for the failure of our earlier ecological prediction.

Why would phage evolve lower productivity? At each MOI, phage from the Unrestricted treatment were significantly more competitive than phage from the Restricted treatment (in competition with a common 'marked' phage strain, see Methods and Fig. 2d). Moreover, a negative correlation between phage productivity and competitive ability was found at all three MOI levels and was significant in two cases (Kendall's test: low MOI  $\tau = -0.25$ ,  $P = 0.04545$ ; intermediate MOI  $\tau = -0.1411$ ,  $P = 0.2655$ ; high MOI  $\tau = -0.4355$ ,  $P = 0.00033$ ). This suggests that the evolution of higher competitive ability led to lower productivity.

These findings are consistent with a 'tragedy of the commons', a scenario that arises when multiple users exploit the same resource. As



**Figure 1 | Stochastic cellular automata predictions.** The global densities of bacteria and phage on a lattice of  $100 \times 100$  wells with an 'Unrestricted' migratory neighbourhood (a) and a 'Restricted' neighbourhood (only the north/south/east/west wells around a focal well can serve as sources of migration) (b). The migration probability is  $m = 0.45$  in both cases. c, A snapshot of the Restricted neighbourhood lattice from time step 100, showing the clumping of bacteria-filled wells (blue), phage-filled wells (red) and empty wells (white).



**Figure 2 | Ecological and evolutionary results for experimental metapopulations.** a, Average bacterial density and b, average phage density through time for each of the three treatments. c, Productivity and d, competitive ability of evolved phage at three multiplicities of infection. We plot mean  $\pm$  s.e.m. and asterisks denote significance at the  $P = 0.05$  level.

unrestrained users displace restrained competitors, the resource (the 'commons') is over-exploited, lowering overall user productivity (the 'tragedy'). The bacteria in each well are a common resource for resident phage. Prudent use of this resource leads to higher productivity of the phage (for example, if phage lengthen the time spent within a host before lysis, then uninfected hosts may continue to replicate, ensuring extended resource access). As shown by Abedon *et al.*<sup>14</sup>, rapacious phage can out-compete their prudent cousins by limiting future access to the host through their own rapid host consumption. The 'tragedy of the commons' is that rapacious types lower overall productivity as they displace more restrained competitors.

Any rapacious phage mutant should displace its ancestor and lower overall productivity of its well. Given periodic dilution, however, less-productive phage subpopulations run a higher risk of extinction. The success of rapacious phage therefore depends on gaining sufficient access to fresh host to compensate for lower productivity. Such access is ensured by unlimited migration in the Unrestricted treatment. By contrast, spatially restricted migration reduces the probability that phage reach fresh hosts, rendering rapacious subpopulations more prone to extinction through dilution. Consequently, the tragedy of the commons is circumvented at the metapopulation scale in the Restricted treatment.

We claim that phage in our experimental metapopulations followed different evolutionary trajectories because of the pattern of migration. One potential concern is that phage under unrestricted migration may experience more replication owing to greater host

access. Perhaps phage evolving under restricted migration did not have a sufficient number of generations (or 'doublings') for competitive variants to arise through mutation and displace their ancestors. Yet we find that the number of phage generations over the entire run,  $T$ , does not differ between the two treatments ( $\bar{T}_{\text{Restricted}} = 64.7$ ,  $\bar{T}_{\text{Unrestricted}} = 64.6$ ,  $P = 1$ , see Supplementary Information I). Evidently, lower evolved productivity and greater access to hosts in the Unrestricted metapopulation are balanced such that phage in both treatments experience a similar amount of replication. Hence, phage in the Restricted metapopulation had ample replication to evolve higher levels of competitiveness had there been universal selection for it.

We extended our earlier simulation to include two phage types, prudent and rapacious, with a trade-off between productivity and competitive ability (see Supplementary Information V). If sufficiently productive, the rapacious type can displace the prudent type in the Unrestricted treatment (Fig. 3a), but not in the Restricted treatment (Fig. 3b), a result that is robust to the shape of the trade-off (Fig. 3c). Rapacious phage fare better in the Unrestricted treatment because: (1) the increased probability of reaching fresh hosts reduces the likelihood of extinction and (2) the increased probability of mixing phage types favours the rapacious competitor<sup>18,19</sup>. In the Restricted treatment, wells of the less-productive rapacious type 'burn out' before spreading very far and prudent types persist by default.

Keeling<sup>18</sup> constructed a metapopulation model where short-term reproductive capacity of a pathogen trades off with long-term persistence. In agreement with his theoretical predictions, we demonstrated empirically that long-term strategies are favoured when dispersal between subpopulations is restricted. Similar trade-offs appear in the evolution of virulence literature<sup>20–30</sup>, where virulence (defined here as an increase in host mortality due to infection) is assumed to covary with transmission<sup>21,24–28</sup> or competitive ability<sup>18,20,23,28</sup>. In our system, wells can be thought of either as microbial subpopulations or as multi-cellular hosts. From this second perspective, the rapacious strain is more virulent because the 'well host' dies sooner (that is, through dilution a less-productive strain goes extinct earlier), resulting in a trade-off between virulence and competitive ability. Assuming this or similar trade-offs, many researchers have investigated the role of host population structure on the evolution of virulence<sup>2,18,20–22,25–27,29,30</sup>. In our experiment, we control the social network of 'well hosts' through different migration treatments. Both our results and theory<sup>18,21,25,29</sup> suggest that highly connected social networks (for example, our Unrestricted treatment) favour virulence. Thus, network topology (between individuals or subpopulations) may profoundly influence the evolutionary trajectory of host–pathogen interactions.

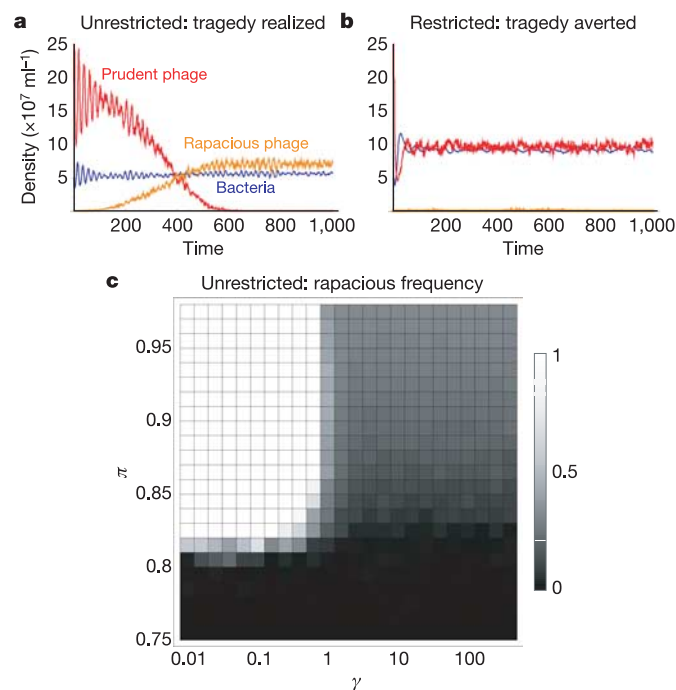
Host–pathogen relationships are just one kind of victim–exploiter interaction, an ubiquitous category that includes prey and their predators and plants and their herbivores. Our results could apply to other victim–exploiter interactions embedded within metapopulations. Specifically, restricted migration between subpopulations may promote the evolution of prudent predation or restrained consumption. In general, restraint in the use of a common resource is a form of cooperation<sup>19</sup>, and we have shown that spatial restrictions in migration within a metapopulation can favour relatively cooperative use of the common resource, thus averting the tragedy of the commons.

## METHODS

Detailed methods and statistical analyses are available in Supplementary Information I.

**Strains and media.** We used the virulent bacteriophage T4 and *E. coli* B strain REL606 (Strep<sup>R</sup> Nov<sup>R</sup> T4<sup>S</sup>), both grown in minimal glucose media supplemented with streptomycin and novobiocin. For the competition assay, we used an rII mutant strain of T4 (ATCC #11303-B40) to which *E. coli* K-12  $\lambda$ -lysogen (ATCC #31278) is resistant, but REL606 is sensitive.

**Well state transitions.** The state of any well can be classified as follows: 'Bacteria' ( $B \approx 3.0 \times 10^7$  cells ml<sup>-1</sup> and  $< 3.0 \times 10^5$  phage ml<sup>-1</sup>), 'Phage 1' ( $P1 > 3.0 \times 10^7$  phage ml<sup>-1</sup>), 'Phage 2' ( $3.0 \times 10^6 < P2 < 3.0 \times 10^7$  phage ml<sup>-1</sup>), 'Phage 3'



**Figure 3 | Evolutionary stochastic cellular automata.** In mixed subpopulations, rapacious phage out-compete prudent cohabitants with probability  $\beta > 0$ , but enter sub-critical levels one dilution earlier with probability  $(1 - \pi) > 0$  ( $\pi$  is a measure of productivity/persistence). Let  $\beta = (1 - \pi)^\gamma$  specify the trade-off between competitive ability and persistence ( $\gamma$  specifies the shape of the trade-off). The fate of rapacious phage (introduced by mutation into a  $100 \times 100$  lattice with bacteria and prudent phage with  $m = 0.45$ ,  $\pi = 0.87$  and  $\gamma = 0.1$ ) is shown given unrestricted (a) and restricted (b) migration. c, The frequency of phage that is rapacious after 3,000 cycles is shown on a greyscale for a number of different parameter combinations (each square is the average rapacious frequency over five simulation runs). With unrestricted migration, rapacious phage that are sufficiently persistent (large  $\pi$ ) completely displace prudent types if the trade-off is sufficiently concave (small  $\gamma$ ). None of the parameter combinations shown allows rapacious phage to persist at high frequencies given restricted migration.



( $3.0 \times 10^5 < P_3 < 3.0 \times 10^6$  phage  $\text{ml}^{-1}$ ), and 'Empty' (E has no bacteria and  $< 3.0 \times 10^5$  phage  $\text{ml}^{-1}$ ). Even with abundant host, phage at subcritical titres (below  $P_3$ ) do not reproduce sufficiently and become extinct over successive dilutions. Bacteria persist in a well with subcritical levels of phage.

We determined a well-state transition in the following manner. A 'focal' well of a given state was diluted tenfold into a well with fresh medium (a 'target' well) and then a 'migrant' well of a given state was diluted approximately tenfold into the target well. After incubation for 12 h, we determined the state of the target well. Consider a table where the rows give the initial state of the focal well and columns give the final state of the target well (Box 1 Table 1). Each entry in this table is the state of the migrant well that yields the relevant row-to-column transition. The entry ' $\emptyset$ ' signifies a row-to-column transition due to dilution alone (that is, no migration). By empirically exploring all five initial states for the focal well and all six migration events (including the 'no migration' event  $\emptyset$ ) we filled the Box 1 Table 1 ( $5 \times 6 = 30$  entries give all possible transitions).

**Experimental treatments.** Each metapopulation consisted of two microtitre plates (192 wells, each with 200  $\mu\text{l}$  growth medium). The initial spatial arrangement of well states was obtained from the 2,000th transfer of a 192-point lattice-based simulation with a Restricted neighbourhood. There were four replicates of each of three treatments: (1) Restricted, (2) Unrestricted and (3) Well-Mixed.

Every 12 h, we used a Biomek FX (Beckman-Coulter) liquid-handling robot with dual-pod action to execute all dilutions, migrations and transfers. Plates were incubated and shaken between transfers (550 r.p.m. on a microtitre shaker at 37 °C). We ran the experiment for a total of 20 transfers. One of the Unrestricted replicates was terminated after 16 transfers owing to contamination.

**Isolating evolved phage strains.** Using a soft-agar overlay with abundant *E. coli* K-12  $\lambda$ -lysogen, we picked four random phage isolates (not rII mutants) from the last transfer of each of the Restricted and Unrestricted replicates (32 isolates in all).

**Bacterial concentration for phage assays.** When phage and bacteria are brought together through migration in the experimental metapopulations, there is an approximately tenfold dilution of both (migration immediately follows dilution). As a consequence, our phage assays are conducted using a tenfold dilution of the bacterial concentration that characterizes a well with fully grown bacteria (state B, see above). This concentration was chosen to mimic the conditions actually faced by the evolving phage.

**Productivity assay.** Each evolved T4 isolate was incubated for 12 h with approximately  $3.0 \times 10^6$  REL606 cells  $\text{ml}^{-1}$  in a microtitre well. The initial and final titres of phage were used to estimate the productivity of the evolved phage (number of progeny produced per parent phage). Each evolved strain was assayed at three MOI levels (initial ratio of phage to bacteria) that mirror levels in the experiment: 'low' (0.05–0.2 MOI), 'intermediate' (0.5–4 MOI) and 'high' (5–60 MOI).

**Competition assay.** We mixed each evolved T4 isolate with rII phage in a 1:10 ratio inside a well with  $3.0 \times 10^6$  REL606 cells  $\text{ml}^{-1}$ . Using the same three MOI levels, we measured phage titres before and after a 12-h incubation period. This was done by diluting and plating the mixture separately in overlays with *E. coli* K-12  $\lambda$ -lysogen (only evolved isolates, not rII, form plaques) and *E. coli* REL606 (both evolved isolates and rII form plaques). The competitive ability of each evolved T4 relative to rII was calculated in one of the following ways:

(1) Per viral replication:

$$v(e, r) = \sqrt[3]{\frac{N_e(12)/N_e(0)}{N_r(12)/N_r(0)}}$$

(2) Per transfer:

$$u(e, r) = \frac{N_e(12)/N_e(0)}{N_r(12)/N_r(0)}$$

(3) Assuming continuous growth:

$$w(e, r) = \frac{\ln[N_e(12)/N_e(0)]}{\ln[N_r(12)/N_r(0)]}$$

where  $N_e(t)$  and  $N_r(t)$  are the densities of evolved T4 and rII, respectively,  $t$  hours after the assay begins. The parameter  $D$  gives the number of viral doublings that occurred over the assay. Where these measures yielded biologically reasonable values, they gave similar results (Fig. 2d uses the 'per viral replication' measure).

Received 12 March; accepted 4 May 2006.

- Bonsall, M. B., French, D. R. & Hassell, M. P. Metapopulation structures affect persistence of predator–prey interactions. *J. Anim. Ecol.* **71**, 1075–1084 (2002).
- Briggs, C. J. & Hoopes, M. F. Stabilizing effects in spatial parasitoid–host and predator–prey models: a review. *Theor. Popul. Biol.* **65**, 299–315 (2004).
- Ellner, S. P. *et al.* Habitat structure and population persistence in an

experimental community. *Nature* **412**, 538–543 (2001).

- Holyoak, M. & Lawler, S. P. Persistence of an extinction-prone predator–prey interaction through metapopulation dynamics. *Ecology* **77**, 1867–1879 (1996).
- Holyoak, M. Habitat patch arrangement and metapopulation persistence of predators and prey. *Am. Nat.* **156**, 378–389 (2000).
- Huffaker, C. B. Experimental studies on predation: dispersion factors and predator–prey oscillations. *Hilgardia* **27**, 343–383 (1958).
- Johst, K. & Schops, K. Persistence and conservation of a consumer–resource metapopulation with local overexploitation of resources. *Biol. Conserv.* **109**, 57–65 (2003).
- Thrall, P. H., Godfree, R. & Burdon, J. J. Influence of spatial structure on pathogen colonization and extinction: a test using an experimental metapopulation. *Plant Pathol.* **52**, 350–361 (2003).
- Hardin, G. The tragedy of the commons. *Science* **162**, 1243–1248 (1968).
- Holyoak, M., Leibold, M. A. & Holt, R. D. *Metacommunities: Spatial Dynamics and Ecological Communities* (Univ. Chicago Press, Chicago, 2005).
- Antonovics, J. in *Ecology, Genetics, and Evolution of Metapopulations* (eds Hanski, I. & Gaggiotti, O. E.) 471–488 (Elsevier, Oxford, 2004).
- Grenfell, B., Bjørnstad, O. N. & Kappey, J. Travelling waves and spatial hierarchies in measles epidemics. *Nature* **414**, 716–723 (2001).
- Keeling, M., Bjørnstad, O. N. & Grenfell, B. in *Ecology, Genetics, and Evolution of Metapopulations* (eds Hanski, I. & Gaggiotti, O. E.) 415–445 (Elsevier, Oxford, 2004).
- Abedon, S. T., Hyman, P. & Thomas, C. Experimental examination of bacteriophage latent-period evolution as a response to bacterial availability. *Appl. Environ. Microbiol.* **69**, 7499–7506 (2003).
- Bohannan, B. J. M. & Lenski, R. E. Linking genetic change to community evolution: insights from studies of bacteria and bacteriophage. *Ecol. Lett.* **3**, 362–377 (2000).
- Lenski, R. E. & Levin, B. R. Constraints on the coevolution of bacteria and virulent phage: a model, some experiments, and predictions for natural communities. *Am. Nat.* **125**, 585–602 (1985).
- Paddison, P. *et al.* The roles of the bacteriophage T4 r genes in lysis inhibition and fine-structure genetics: A new perspective. *Genetics* **148**, 1539–1550 (1998).
- Keeling, M. Evolutionary trade-offs at two time-scales: competition versus persistence. *Proc. R. Soc. Lond. B* **267**, 385–391 (2000).
- Kreft, J. U. Biofilms promote altruism. *Microbiology* **150**, 2751–2760 (2004).
- Bergstrom, C. T., McElhany, P. & Real, L. A. Transmission bottlenecks as determinants of virulence in rapidly evolving pathogens. *Proc. Natl Acad. Sci. USA* **96**, 5095–5100 (1999).
- Boots, M. & Sasaki, A. 'Small worlds' and the evolution of virulence: infection occurs locally and at a distance. *Proc. R. Soc. Lond. B* **266**, 1933–1938 (1999).
- Bull, J. J., Molineux, I. J. & Rice, W. R. Selection of benevolence in a host–parasite system. *Evolution* **45**, 875–882 (1991).
- Cooper, V. S. *et al.* Timing of transmission and the evolution of virulence of an insect virus. *Proc. R. Soc. Lond. B* **269**, 1161–1165 (2002).
- Ebert, D. Virulence and local adaptation of a horizontally transmitted parasite. *Science* **265**, 1084–1086 (1994).
- Galvani, A. P. Epidemiology meets evolutionary ecology. *Trends Ecol. Evol.* **18**, 132–139 (2003).
- Herre, E. A. Population structure and the evolution of virulence in nematode parasites of fig wasps. *Science* **259**, 1442–1445 (1993).
- Lipsitch, M., Herre, E. A. & Nowak, M. A. Host population structure and the evolution of virulence: a 'law of diminishing returns'. *Evolution* **49**, 743–748 (1995).
- Nowak, M. A. & May, R. M. Superinfection and the evolution of parasite virulence. *Proc. R. Soc. Lond. B* **255**, 81–89 (1994).
- O'Keefe, K. J. & Antonovics, J. Playing by different rules: The evolution of virulence in sterilizing pathogens. *Am. Nat.* **159**, 597–605 (2002).
- Thrall, P. H. & Burdon, J. J. Evolution of virulence in a plant host–pathogen metapopulation. *Science* **299**, 1735–1737 (2003).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank Y. Dang for help in the laboratory and the BioTechnology Resource Center at the University of Minnesota for robot access. We thank S. Abedon, C. Bergstrom, J. Bull, J. Fletcher, K. Koelle, C. Lehman, B. Levin and D. Stephens for useful feedback on this project and manuscript. This work was partially supported by an NSF grant to C.N. and an NIH grant to A.M.D.

**Author Contributions** B.K., A.M.D. and B.J.M.B. designed the experiments. B.K. and A.M.D. worked out the robotic protocols. B.K. programmed the robot, executed the experiments, and conducted the assays. B.K. and C.N. coded and analysed the empirically calibrated and evolutionary models. B.K., C.N. and A.M.D. conducted the statistical analysis. All authors contributed to the writing of the manuscript.

**Author Information** Reprints and permissions information is available at [npg.nature.com/reprintsandpermissions](http://npg.nature.com/reprintsandpermissions). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to B.K. ([kerrb@u.washington.edu](mailto:kerrb@u.washington.edu)).

# Retroviral invasion of the koala genome

Rachael E. Tarlinton<sup>1</sup>, Joanne Meers<sup>2</sup> & Paul R. Young<sup>1</sup>

Endogenous retroviruses are a common ancestral feature of mammalian genomes with most having been inactivated over time through mutation and deletion<sup>1</sup>. A group of more intact endogenous retroviruses are considered to have entered the genomes of some species more recently, through infection by exogenous viruses<sup>2</sup>, but this event has never been directly proved. We have previously reported koala retrovirus (KoRV) to be a functional virus that is associated with neoplasia<sup>3</sup>. Here we show that KoRV also shows features of a recently inserted endogenous retrovirus that is vertically transmitted. The finding that some isolated koala populations have not yet incorporated KoRV into their genomes, combined with its high level of activity and variability in individual koalas, suggests that KoRV is a virus in transition between an exogenous and endogenous element. This ongoing dynamic interaction with a wild species provides an exciting opportunity to study the process and consequences of retroviral endogenization in action, and is an attractive model for studying the evolutionary event in which a retrovirus invades a mammalian genome.

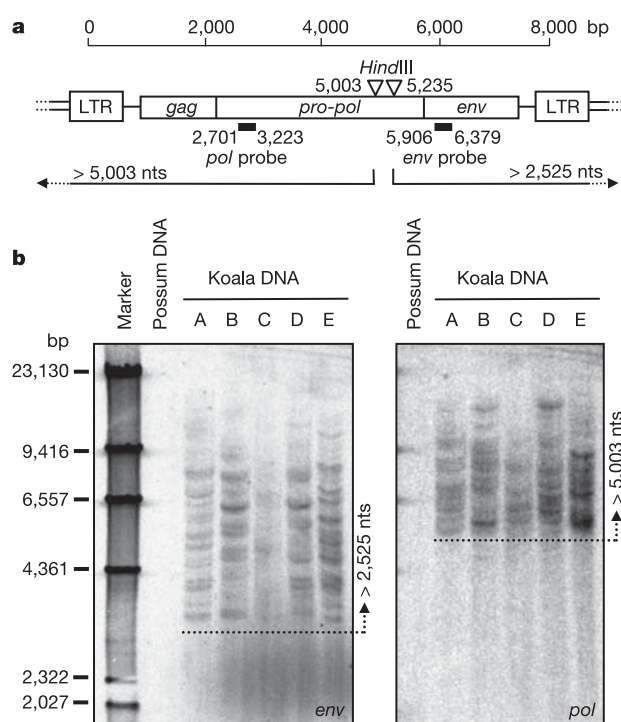
The koala retrovirus, KoRV, was originally identified as an endogenous retrovirus based on its presence in all koalas tested<sup>4</sup>. However, it was unusual in that it possessed a full-length replication-competent genome, viral particles could be produced from cultured koala peripheral blood mononuclear cells (PBMCs), and its sequence indicated a close relationship with gibbon ape leukaemia virus (GALV)—a pathogenic exogenous virus of gibbons<sup>5</sup>. KoRV is also actively transcribed in koalas, and both haematopoietic neoplasia and chlamydiosis are positively linked with plasma viral load<sup>3</sup>.

We have found previously that the KoRV genome copy number varies between individual animals<sup>3</sup>. To examine this apparently ongoing dynamic interaction between KoRV and the koala population, we examined variation in KoRV inserts between individual koalas. Southern blotting was performed on *Hind*III-digested DNA from lymph node tissue from unrelated animals from captive and wild populations (Fig. 1 and Supplementary Table 1). Hybridization of probes derived from the envelope (*env*) (Fig. 1b, left panel) and polymerase (*pol*) (Fig. 1b, right panel) regions of KoRV revealed considerable variation in both the number and pattern of integrations. Only a few bands were common to some animals, with none common to all. This pattern is not typical for an endogenous virus of a single host species where a higher degree of conservation of bands could be expected<sup>6,7</sup>. Cytogenetic analysis of koala PBMCs with a KoRV probe was also performed (Supplementary Fig. 1a), which showed that inserts were spread throughout the koala genome, unlike the pattern seen for the only other marsupial retrovirus studied in detail—the kangaroo endogenous retrovirus KeRV-1 (ref. 8).

The smallest fragments revealed by Southern blotting (Fig. 1b) were no less than 5,000 (*pol* probe) and 2,500 (*env* probe) nucleotides in length, suggesting that the majority of KoRV inserts are full length, although polymerase chain reaction (PCR) data from a previous report<sup>4</sup>, and our own unpublished work, demonstrates that truncated inserts do occur in some koalas. Further evidence for KoRV inserts

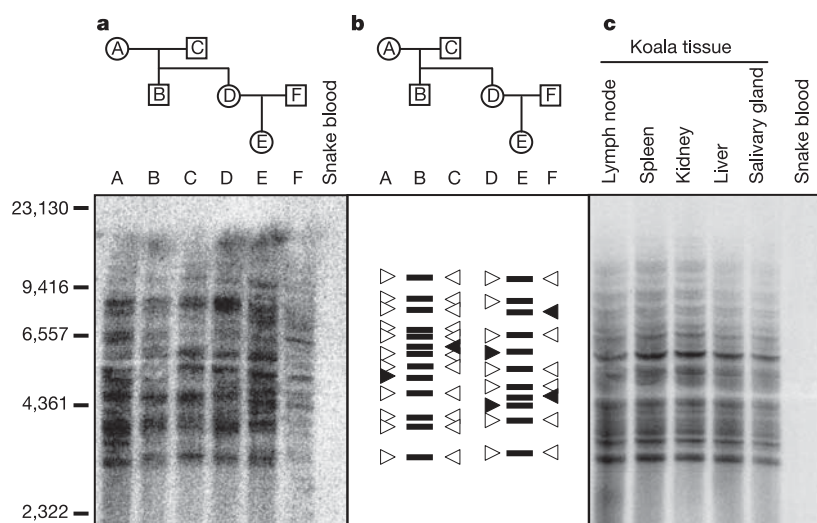
retaining functional activity, apart from their apparent full length, was suggested by sequencing of the hypervariable region of the KoRV envelope gene (determined by comparison with GALV<sup>9</sup>; data not shown). Considerable variation was seen both between and within animals; however, 63.5% of these were silent mutations and no premature stop codons were found. Deletions of up to nine base pairs in the envelope gene region were observed, but all deletions were found to be in frame. Important structural features, such as the eight histidine residues conserved among gamma-retroviruses, were maintained<sup>10</sup>. This type of selection for functional quasi-species has been reported for exogenous viruses<sup>11,12</sup>, but is not typical for endogenous viruses where loss of functionality is expected to occur through random mutation<sup>13</sup>.

On the basis of these data, KoRV seemed more like an exogenous, rather than endogenous, virus. Although no standard definition of



**Figure 1 | Patterns of insertion of KoRV in the koala genome.** **a**, Schematic showing full-length KoRV in the koala genome, the location of *Hind*III sites, and the *pol* and *env* probes. Expected minimum sizes of DNA fragments of *Hind*III-cut full-length inserts are shown below. LTR, long terminal repeat; bp, base pairs; nts, nucleotides. **b**, Southern blot analysis of *Hind*III-digested DNA from lymph nodes of unrelated koalas (A–E) hybridized with either the *env* (left panel) or *pol* (right panel) probes. Possum DNA was run as a control. The minimum size of KoRV fragments corresponds to that expected for full-length inserts. Marker track is *Hind*III-digested  $\lambda$  phage DNA.

<sup>1</sup>School of Molecular and Microbial Sciences, and <sup>2</sup>School of Veterinary Science, University of Queensland, St Lucia 4072, Queensland, Australia.



**Figure 2 | Patterns of KoRV genome insertions are inherited.** **a**, Southern blot of *Hind*III-digested DNA from blood of related koalas hybridized with the KoRV *env* probe. The family lineage of animals A–F is represented above the Southern blot. **b**, Representation of the insert profiles of offspring B and E with inheritance from parents indicated by open (either parent) or

filled (single parent) arrowheads. **c**, Southern blot of *Hind*III-digested DNA from different tissues of an individual koala. DNA from snake blood was run as a control for both blots. Marker track is *Hind*III-digested  $\lambda$  phage DNA.

an endogenous virus exists, most sources agree that they must be integrated into germ-line tissue and passed on to offspring as a mendelian allele<sup>2</sup>. Two methods were used to confirm KoRV integration into germ-line tissue. Single-cell fluorescent PCR of koala sperm was performed in order to eliminate the possibility of lysed blood or somatic cell contamination yielding false PCR positives, and this confirmed that KoRV was present in koala germ cells (Supplementary Fig. 1c). This result was supported by fluorescence *in situ* hybridization (FISH) analysis of koala whole sperm with a KoRV-specific probe. (Supplementary Fig. 1d).

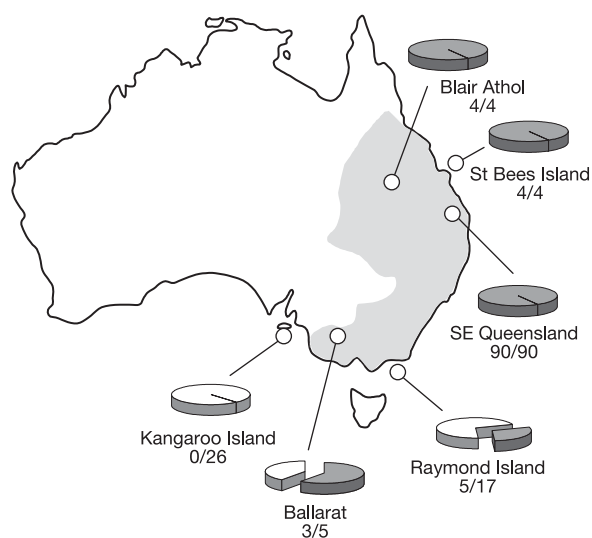
Given the presence of KoRV in germ-line cells, we investigated family lineages for inheritance of specific insert patterns. Southern blotting of *Hind*III-digested DNA extracted from blood cells of related animals from a captive colony and hybridization with the KoRV *env* probe (Fig. 2a) demonstrated that these animals showed a higher degree of insert pattern similarity than unrelated animals (compare with Fig. 1b). Banding patterns in the offspring (Fig. 2b) were found to be a combination of parental patterns, with some bands being inherited across three generations. The viral insert pattern did not vary across tissues in an individual, further indicating that the virus was endogenous (Fig. 2c).

To examine KoRV insertions in the koala population as a whole, blood samples were obtained from wild koalas over the geographical distribution of the species, and KoRV *pol*- and *env*-specific PCR and real-time PCR performed on extracted DNA. Unexpectedly, KoRV was not detected in any of the animals from Kangaroo Island (situated off the coast of South Australia; Fig. 3). Other populations in southern Australia showed a mixed prevalence of KoRV. This was in contrast to the uniform detection of KoRV in Queensland animals (Fig. 3), suggesting ongoing infection and endogenization in the koala population, spreading from northern to southern Australia. Given the very high levels of KoRV activity, both endogenous and exogenous spread of the virus may be occurring in areas of mixed KoRV status. In support of our earlier observations of a link between KoRV and disease in koalas<sup>3</sup>, haematology data from a survey of wild Kangaroo Island animals showed no cases of leukaemia in 225 animals investigated. In contrast, among KoRV-infected animals from southeastern Queensland, 3 out of 63 succumbed to neoplastic disease over the course of our 18-month study<sup>3</sup>.

Koalas were largely exterminated on mainland southern Australia as a result of hunting during the late nineteenth and early twentieth

centuries. Koala populations were established on a small number of islands (including Kangaroo Island) in the early 1900s and have remained isolated since the 1920s. These populations have since been used to restock the mainland in ongoing relocation programmes<sup>14,15</sup>. The fact that these island and other southern Australian populations have such a low prevalence of KoRV suggests that the virus was not present in the foundation stock in the early 1900s, and that the virus has entered the koala population only in this very recent time frame.

One of the puzzling aspects of KoRV when it was first isolated was the very high degree of similarity with GALV. Speculation that the two viruses diverged only recently<sup>16</sup> is supported by our study and the search for a potential common ancestor is currently underway. Given these observations, KoRV seems to have entered the koala population



**Figure 3 | Prevalence of KoRV in geographically distinct koala populations.** Distribution of wild koala populations in Australia is shown by the grey shading<sup>21</sup>. Prevalence rates of KoRV in different study populations were defined by PCR and real-time PCR results on DNA extracted from blood samples. Data are represented as pie graphs with positive (dark grey) or negative (white) results shown—values indicate number of positive samples/number of samples tested.



within the last 100 years and to still be invading the koala genome. This is in marked contrast to other modern endogenous retroviruses where the most recent estimate for endogenization is a porcine virus that entered its host approximately 5,000 years ago<sup>17</sup>. Although there are functional endogenous proviral loci in other species, and indeed endogenization of single proviruses derived from pre-existing loci have been observed in mice<sup>2,18–20</sup>, KoRV is unique in that we are observing the initial entry of a new family of proviruses into a wild host genome.

Although the association of KoRV with disease in a vulnerable species raises important management issues for both wild and captive koalas, the dynamic interaction between this virus and its new host also provides an exciting opportunity to study the process of endogenization in action in a natural setting. This may help shed light on an almost universal process in species development, and answer fundamental questions as to the establishment and role of endogenous retroviruses in their hosts.

## METHODS

**Southern blotting.** DNA was isolated from whole blood or tissue samples using the Qiagen DNA Mini or Qlump DNA mini kit (Qiagen). Southern blots of samples consisting of 10 µg of DNA digested with *Hind*III were hybridized with <sup>32</sup>P-dCTP-labelled KoRV *pol* or *env* probes (described in Supplementary Methods). Detection was performed on a phosphorimager plate (Kodak) and SI phosphorimager (Molecular Dynamics).

**Cytogenetics/fluorescence *in situ* hybridization.** Koala PBMCs were isolated from whole blood and cultured as described in Supplementary Methods. Cells were harvested after three days and chromosomes prepared using standard methods with 3 h of incubation with colcemid. Sperm for FISH analysis were prepared as described in Supplementary Methods. Digoxigenin (DIG)-labelled KoRV probes were prepared as described in Supplementary Methods. Hybridization was performed using the DIG hybridization kit (Roche) and images were collected with an Axiophot imaging system (Zeiss).

**PCR.** DNA was isolated from whole koala blood or sperm using the Qlump DNA mini kit (Qiagen). Real-time PCR was performed as described previously<sup>3</sup>. Additional PCR primer sets used were: envelope gene primers, Env2 5'-ATGCTTCTCATCTCAAACCG-3' (sense) and 5'-TTCAAGTCTCCGACC TCC-3' (antisense), Env3 5'-GGAGGTGCGGAGGACTTGAA-3' (sense) and 5'-CGTGGTGGGAGATGGAGTAC-3' (antisense); and polymerase gene primers, 5'-CCTTGGACCAAGAGACTTTGA-3' (sense) and 5'-TCAAATCTTGG ACTGGCCGA-3' (antisense). β-actin PCR was performed in parallel with KoRV PCR as a control for DNA quality. Primers were: 5'-AGATCATTGCCCC ACCT-3' (sense) and 5'-TGGAAGGCCAGATTC-3' (antisense). Cycling parameters are described in Supplementary Methods.

**Single-cell sperm PCR.** Koala sperm were obtained by epididymal washing. Individual sperm were isolated and both sperm and negative controls were treated with Promega DNase as described in Supplementary Methods. Primers for multiplex fluorescent PCR were: KoRVmgbf, FAM-5'-TTGGAGGAGGAA TACCGATTACAC-3' and KoRVmgb, 5'-GCCAGTCCCATACCTGCCTT-3' (Qiagen). PCR cycling and analysis by capillary electrophoresis are described in Supplementary Methods.

**Permits.** The work in this paper was conducted under animal ethics approvals issued by the University of Queensland Animal Ethics Committee and the South Australian Department of the Environment and Heritage, Wildlife Ethics Committee. Research permits for wildlife research were obtained from New South Wales National Parks and Wildlife, Victorian Department of Sustainability and the Environment, and South Australian National Parks and Wildlife.

Received 4 April; accepted 26 April 2006.

1. Taruscio, D. & Mantovani, A. Factors regulating endogenous retroviral sequences in human and mouse. *Cytogenet. Genome Res.* **105**, 2–4 (2004).
2. Boeke, J. D. & Stoye, J. P. in *Retroviruses* (eds Coffin, J. M., Hughes, S. H. & Varmus, H. E.) 343–346 (Cold Spring Harbor Laboratory Press, New York, 1997).
3. Tarlinton, R., Meers, J., Hanger, J. & Young, P. Real-time reverse transcriptase

PCR for the endogenous koala retrovirus reveals an association between plasma viral load and neoplastic disease in koalas. *J. Gen. Virol.* **86**, 783–787 (2005).

4. Hanger, J. J., Bromham, L. D., McKee, J. J., O'Brien, T. M. & Robinson, W. F. The nucleotide sequence of koala (*Phascogalea cinereus*) retrovirus: a novel type C endogenous virus related to gibbon ape leukemia virus. *J. Virol.* **74**, 4264–4272 (2000).
5. Herniou, E. *et al.* Retroviral diversity and distribution in vertebrates. *J. Virol.* **72**, 5955–5966 (1998).
6. Resnick, R. M., Boyce-Jacino, M. T., Fu, Q. & Faras, A. J. Phylogenetic distribution of the novel avian endogenous provirus family EAV-O. *J. Virol.* **64**, 4640–4653 (1990).
7. Carlson, J. *et al.* Chromosomal distribution of endogenous Jaagsiekte sheep retrovirus proviral sequences in the sheep genome. *J. Virol.* **77**, 9662–9668 (2003).
8. Waugh O'Neill, R. J., O'Neill, M. J. & Graves, J. A. Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. *Nature* **396**, 68–72 (1998).
9. Delassus, S., Sonigo, P. & Wain-Hobson, S. Genetic organisation of gibbon ape leukemia virus. *Virology* **173**, 205–213 (1989).
10. Battini, J., Heard, J. M. & Danos, O. Receptor choice determinants in the envelope glycoproteins of amphotropic, xenotropic, and polytropic murine leukemia viruses. *J. Virol.* **66**, 1468–1475 (1992).
11. Tailor, A. K. & Willet, B. J. in *The 15th International Workshop on Retroviral Pathogenesis* abstr. B.5 (Glasgow, Scotland, 2003).
12. Fauci, A. S. & Desrosiers, R. C. in *Retroviruses* (eds Coffin, J. M., Hughes, S. H. & Varmus, H. E.) 587–636 (Cold Spring Harbor Laboratory Press, New York, 1997).
13. Belshaw, R. *et al.* Long-term reinfection of the human genome by endogenous retroviruses. *Proc. Natl Acad. Sci. USA* **101**, 4894–4899 (2004).
14. Martin, R. & Handasyde, K. in *Managing Koalas in New South Wales: Proceedings of the Koala Summit held at the University of Sydney 7–8 November 1988* (eds Lunney, D., Urquhart, C. A. & Reed, P.) 58–66 (NSW National Parks and Wildlife Service, Sydney, 1990).
15. Masters, P., Duka, T., Berris, S. & Moss, G. Koalas on Kangaroo Island: from introduction to pest status in less than a century. *Wildl. Res.* **31**, 267–272 (2004).
16. Bromham, L. The human zoo: endogenous retroviruses in the human genome. *Trends Ecol. Evol.* **17**, 91–97 (2002).
17. Mang, R., Maas, J., Chen, X., Goudsmit, J. & Van der Kuyl, A. Identification of a novel type C porcine endogenous retrovirus: evidence that copy number of endogenous retroviruses increases during host inbreeding. *J. Gen. Virol.* **82**, 1829–1834 (2001).
18. Jaenisch, R. Germ line integration and Mendelian transmission of the exogenous Moloney leukemia virus. *Proc. Natl Acad. Sci. USA* **73**, 1260–1264 (1976).
19. Frankel, W. N., Stoye, J. P., Taylor, B. A. & Coffin, J. M. Genetic identification of endogenous polytropic proviruses by using recombinant inbred mice. *J. Virol.* **63**, 3810–3821 (1989).
20. Steffen, D. L., Taylor, B. A. & Weinberg, R. A. Continuing germ line integration of AKC proviruses during the breeding of AKR mice and derivative recombinant inbred strains. *J. Virol.* **42**, 165–175 (1982).
21. Sharp, A. *The Koala Book* (David Bateman, Auckland, 1995).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank J. Hanger, J. McKee, D. Hume and M. Eiden for discussions; Dreamworld Themepark, Queensland Parks and Wildlife Service, The South Australian Department of the Environment and Heritage, D. Higgins, S. Johnston and W. Ellis for access to, and supply of, samples; D. Irwin and staff at the Australian Genome and Research Facility for technical assistance with the single-cell PCR; B. Carroll and staff in the School of Molecular and Microbial Sciences, University of Queensland, for technical assistance with the Southern blotting analyses. This work was supported by Dreamworld Themepark and the Australian Research Council.

**Author Contributions** R.E.T. carried out all experiments shown in Figs 1–3 and in Supplementary Information. All authors were involved in the conceptual design and interpretation of the experimental work with direction and supervision provided by J.M. and P.R.Y. All authors participated in the drafting of the manuscript.

**Author Information** Reprints and permissions information is available at [npg.nature.com/reprintsandpermissions](http://npg.nature.com/reprintsandpermissions). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.R.Y. (p.young@uq.edu.au).

## LETTERS

# Anti-apoptotic function of a microRNA encoded by the HSV-1 latency-associated transcript

A. Gupta<sup>1</sup>, J. J. Gartner<sup>1</sup>, P. Sethupathy<sup>2</sup>, A. G. Hatzigeorgiou<sup>2</sup> & N. W. Fraser<sup>1</sup>

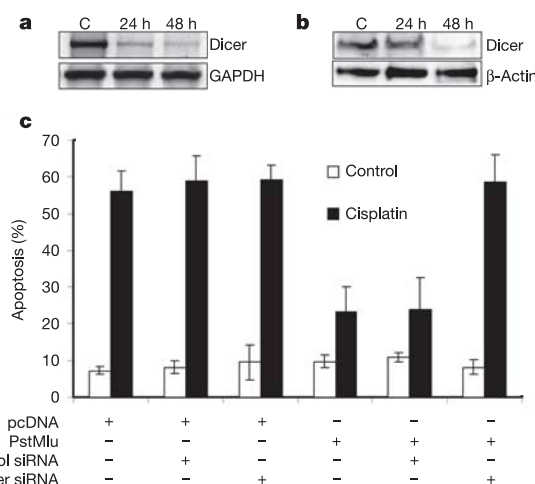
MicroRNAs (miRNAs) are a class of small RNA molecules that regulate the stability or the translational efficiency of target messenger RNAs (mRNAs)<sup>1,2</sup>. The latency-associated transcript (LAT) of herpes simplex virus-1 (HSV-1) is the only viral gene expressed during latent infection in neurons<sup>3</sup>. LAT inhibits apoptosis and maintains latency by promoting the survival of infected neurons<sup>4</sup>. No protein product has been attributed to the LAT gene and the mechanism by which LAT protects cells from apoptosis is not yet known. Here we show that a miRNA encoded by the HSV-1 LAT gene confers resistance to apoptosis. Neuroblastoma cells transfected with a fragment of the LAT gene show reduced susceptibility to cell death. The anti-apoptotic function of LAT has been mapped to a region within the first exon<sup>5,6</sup>. We have identified and characterized a microRNA (miR-LAT) generated from the exon 1 region of the HSV-1 LAT gene. The LAT miRNA was found to accumulate in cells transiently transfected with the LAT gene fragment or infected with a wild-type strain of HSV-1. A mutant virus in which a 372-nucleotide fragment encompassing the mature miRNA was deleted neither protected the infected cells from apoptosis nor generated an miRNA. miR-LAT exerts its anti-apoptotic effect by downregulation of transforming growth factor (TGF)- $\beta$  1 and SMAD3 expression, both of which are functionally linked in the TGF- $\beta$  pathway. Our results suggest that the miRNA encoded by the HSV-1 LAT gene regulates the induction of apoptosis in infected cells by modulation of TGF- $\beta$  signalling and thus contributes to the persistence of HSV in a latent form in sensory neurons.

First we confirmed the anti-apoptotic function of LAT. As expected, a 2.9-kilobase (kb) fragment of the HSV-1 LAT gene (pcDNA-PstMlu) reduced susceptibility to stress-induced apoptosis in SY5Y and HeLa cells (Supplementary Fig. 1). Although it has been shown that LAT promotes neuronal survival after HSV infection by reducing apoptosis<sup>4</sup>, the mechanism by which it regulates this process is still not understood. We hypothesized that the HSV-1 LAT gene exerts its anti-apoptotic effect by encoding a miRNA. To test this hypothesis, we attenuated the function of the RNase III enzyme Dicer by using specific short interfering RNA (siRNA) (Fig. 1a, b) and evaluated its effect on LAT-mediated resistance to apoptosis. Dicer-targeted siRNA abrogated the ability of a short hairpin RNA (shRNA) directed against green fluorescent protein (GFP-shRNA)<sup>7</sup> to block GFP expression in SY5Y cells (Supplementary Fig. 2). Dicer-targeted siRNA significantly inhibited the anti-apoptotic effect of LAT as compared to cells transfected with a control nonspecific siRNA (Fig. 1c).

Computational analysis using RNAMOT<sup>8</sup> and MFOLD<sup>9</sup> revealed the presence of a sequence within exon 1 of HSV-1 LAT containing a hairpin loop reminiscent of precursor miRNAs (Fig. 2a). Next we cloned the small RNAs from SY5Y cells transfected with pcDNA-PstMlu. Colony hybridization with the Sty-Sty region of LAT revealed that 7% of the clones contained miR-LAT sequences. Sequencing of

positive clones confirmed the predicted sequence of miR-LAT shown in Fig. 2a. We cloned miR-LAT multiple times. No other HSV-1 miRNA was present in the sequenced clones. The sequences present in the HSV-1 LAT region encoding miR-LAT are conserved in 17+, F and McKrae strains of HSV-1 (Fig. 2a). Transfection of cells with a LAT deletion construct ( $\Delta$ Sty)—deletion of a 372-nucleotide fragment encompassing the predicted mature miRNA (Fig. 2a)—showed significant reduction in protection from cisplatin-induced apoptosis (Fig. 2b).

A synthetic RNA oligonucleotide corresponding to pre-miR-LAT was processed *in vitro* by a recombinant Dicer enzyme to yield a ~20-nucleotide product (Fig. 2c). Next we performed a northern blot analysis to evaluate the expression of miR-LAT in LAT-expressing cells. Only one of the strands of the double-stranded hairpin precursor is energetically favoured to enter the RISC complex and accumulate in the cell as the mature miRNA<sup>10</sup>. Therefore we used two probes: the 5' probe comprising of the 5' arm of the hairpin, and the 3' probe comprising of the 3' arm of the hairpin. The 3' probe revealed a ~60-nucleotide and a ~20-nucleotide band (Fig. 2d), corresponding to pre-miRNA and mature miRNA, respectively. The 5' probe did not show any mature miRNA (Fig. 2d); however, a ~60-nucleotide pre-miRNA band was detected by the 5' probe. This indicates that



**Figure 1 | LAT inhibits apoptosis by a Dicer-dependent mechanism.**

**a**, RT-PCR analysis of the transcript levels of Dicer and GAPDH after transfection of SY5Y cells with siRNA targeting Dicer. **b**, Western blot analysis for Dicer and  $\beta$ -actin using whole-cell lysates of SY5Y cells transfected with siRNA targeting Dicer. **c**, SY5Y cells were transfected with the indicated plasmids, treated with cisplatin and the percentage of apoptosis was determined. Data represent the mean of three independent experiments  $\pm$  s.d.

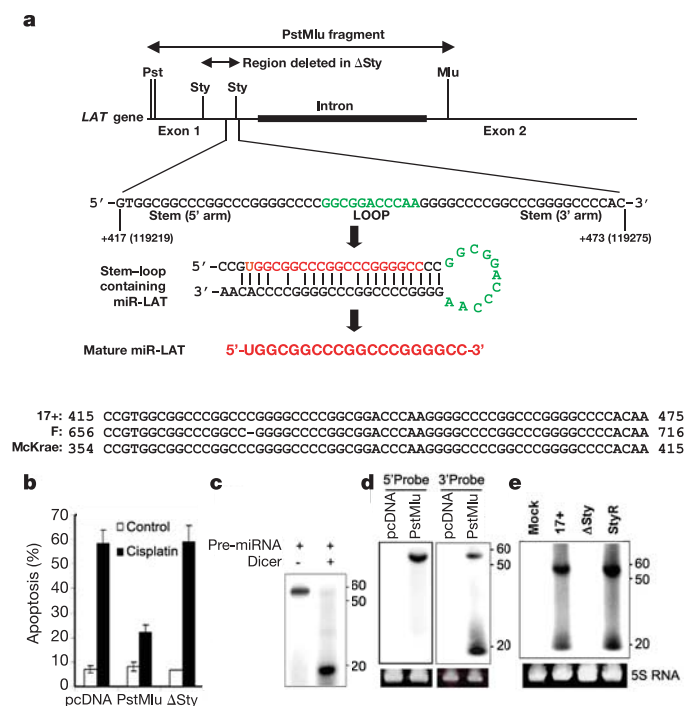
<sup>1</sup>Department of Microbiology, University of Pennsylvania School of Medicine, and <sup>2</sup>Department of Genetics and Penn Center for Bioinformatics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA.

only the 5' arm of the pre-miRNA hairpin constitutes the mature miR-LAT. Cells transfected with the plasmid pcDNA- $\Delta$ Sty, a deletion mutant of *LAT*, were not protected from apoptosis (Fig. 2b) and northern blot analysis did not detect any pre-miRNA or mature miRNA (data not shown). Next we determined the generation of miR-LAT in virus-infected cells. For this purpose we used the wild-type virus (17+), mutant virus ( $\Delta$ Sty) with a 372-nucleotide deletion in exon 1 (position +76 to +447) of the *LAT* gene, and a  $\Delta$ Sty rescued virus (StyR). The mutation did not affect the ability of the virus to establish infection in tissue culture or the expression of other viral genes (data not shown). As shown in Fig. 2e, ~20-nucleotide mature miR-LAT was detected in cells infected with wild-type and StyR virus. No mature miR-LAT was found in cells infected with the mutant virus (Fig. 2e). Furthermore, the  $\Delta$ Sty virus was less efficient in protecting cells from stress-induced apoptosis as compared to the wild-type and StyR virus (Fig. 3a, b; see also Supplementary Fig. 3).

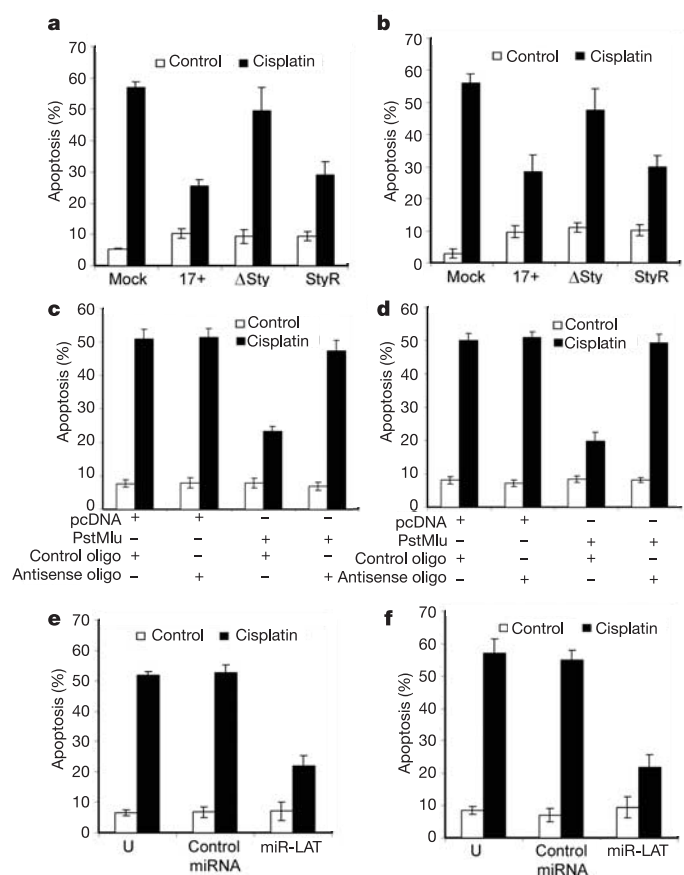
To determine the role of miR-LAT in providing resistance to apoptosis we used 2'-O-methyl antisense oligonucleotides specific for miR-LAT. As shown in Fig. 3c, d, co-transfection of the 2'-O-methyl antisense oligonucleotide with the *LAT* fragment (pcDNA-PstMlu) significantly attenuated its anti-apoptotic effect, whereas a control 2'-O-methyl oligonucleotide had no such effect. Both antisense and control 2'-O-methyl oligonucleotides did not show any effect on apoptosis when co-transfected with control (pcDNA) plasmid (Fig. 3c, d; see also Supplementary Fig. 4). To elucidate further the anti-apoptotic function of miR-LAT we transfected cells with a synthetic double-stranded RNA oligonucleotide corresponding to mature miR-LAT. This synthetic miR-LAT oligonucleotide

caused significant reduction in apoptosis as compared to control miRNA (Fig. 3e, f; see also Supplementary Fig. 5).

To gain an insight into the mechanism by which miR-LAT regulates apoptosis, we sought to identify targets of miR-LAT. Computational analysis using miRanda<sup>11</sup> identified transforming growth factor- $\beta$ 1 (TGF- $\beta$ ) and SMAD3 as targets of miR-LAT, both of which are functionally related in TGF- $\beta$  signalling. TGF- $\beta$  is a potent inhibitor of cell growth and an inducer of apoptosis<sup>12</sup>. SMAD2 and SMAD3 are phosphorylated by the activated TGF- $\beta$  receptor, form complexes with SMAD4, and together accumulate in the nucleus to regulate transcription of target genes that have important roles in diverse cellular processes<sup>13</sup>. TGF- $\beta$ -induced cell death is associated with changes in the expression, localization and activation of both pro- and anti-apoptotic members of the Bcl2 family, as well as activation of caspases<sup>14</sup>. The 3' untranslated region (UTR) of TGF- $\beta$  and SMAD3 contain sequences with partial homology to miR-LAT (Fig. 4a). The miR-LAT response element in the 3'UTR of TGF- $\beta$  is evolutionarily conserved (Fig. 4a). Given that miRNAs function in RNA silencing pathways either by targeting mRNAs for degradation or by repressing translation, we determined the effect of HSV-1 *LAT* on mRNA and protein levels of TGF- $\beta$  and SMAD3. We observed a substantial decrease in the mRNA levels of TGF- $\beta$  and SMAD3 in cells transfected with pcDNA-PstMlu or infected with



**Figure 2 | The *LAT* gene of HSV-1 codes for a miRNA.** **a**, Schematic representation of the HSV-1 *LAT* gene. Sequence of mature miRNA is shown in red. Bottom panel: sequence conservation of miR-LAT-encoding region in 17+, F and McKrae strains of HSV. **b**, SY5Y cells were treated with cisplatin 24 h after transfection and the percentage of apoptosis was determined. Data represent the mean of three independent experiments  $\pm$  s.d. **c**, Pre-miR-LAT is processed *in vitro* by recombinant Dicer. **d**, Northern blot analysis of the *LAT* miRNA in SY5Y cells transfected with pcDNA or pcDNA-PstMlu. **e**, Northern blot detection of *LAT* miRNA in cells infected with the indicated viruses. An ethidium-bromide-stained picture of 5S RNA is shown as a loading control. The migration of a DNA oligonucleotide ladder (bp) is indicated.



**Figure 3 | The *LAT* region of HSV-1 17+ protects cells from apoptosis.** **a**, **b**, SY5Y (**a**) and HeLa (**b**) cells were infected as indicated and treated with cisplatin 16 h after infection. Data represent the mean of three independent experiments  $\pm$  s.d. **c**, **d**, SY5Y (**c**) and HeLa (**d**) cells were transfected as indicated, treated with cisplatin 24 h after transfection, and the percentage of apoptosis was determined. Data represent the mean of three independent experiments  $\pm$  s.d. **e**, **f**, SY5Y (**e**) and HeLa (**f**) cells were transfected with the indicated miRNA, and the percentage of apoptosis was determined. Data represent the mean of three independent experiments  $\pm$  s.d. U, untransfected.



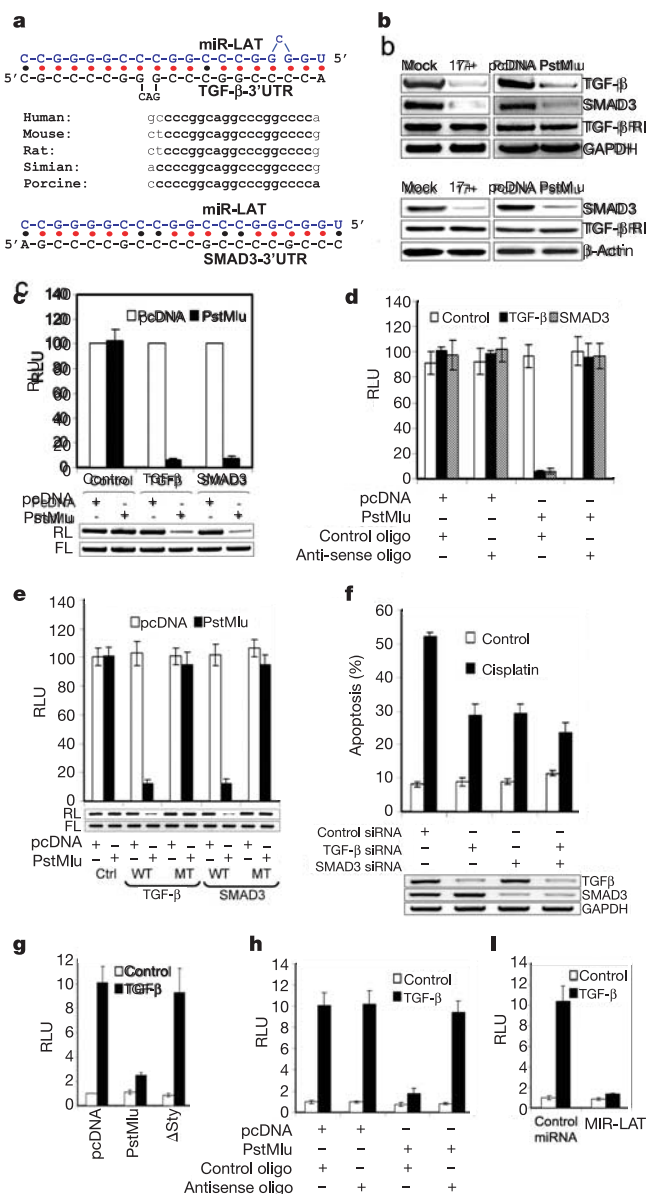
wild-type 17+ virus (Fig. 4b), which was accompanied by a concomitant decrease in endogenous SMAD3 (Fig. 4b) and TGF- $\beta$  protein (Supplementary Fig. 6). We observed a significant decrease in the activity of the *Renilla* luciferase reporter gene fused to the full length 3' UTRs of TGF- $\beta$  or SMAD3 in *LAT*-expressing SY5Y cells, which was accompanied by a decrease in *Renilla* luciferase mRNA (Fig. 4c). There was no effect on the activity and mRNA levels of a *Renilla* luciferase reporter gene lacking the miR-*LAT* response elements (Fig. 4c). Furthermore, *LAT* expression had no effect on mRNA levels of firefly luciferase, which was used as a transfection control (Fig. 4c). *LAT* expression had no effect on the activity of a *Renilla* luciferase reporter construct having the 3' UTR of TGF- $\beta$  or SMAD3 in the presence of a 2'-*O*-methyl antisense oligonucleotide specific for miR-*LAT* (Fig. 4d). To confirm that repression of TGF- $\beta$  and SMAD3 by miR-*LAT* is via the predicted miR-*LAT*-binding site in their 3' UTRs (Fig. 4a), we cloned double-stranded oligonucleotides comprising the wild-type target site or a mutant target site that disrupts miR-*LAT* binding at the 3' end of the *Renilla* luciferase gene (Supplementary Fig. 7). These clones contained a single copy of either the wild-type or the mutated target site. The wild-type construct (TGF- $\beta$  and SMAD3) showed up to a 90% reduction in luciferase activity, which was accompanied by a decrease in *Renilla*

luciferase mRNA levels (Fig. 4e). The mutant construct did not significantly decrease luciferase activity or mRNA levels of a *Renilla* luciferase reporter gene (Fig. 4e). Thus, the proposed site (Fig. 4a) is the major miR-LAT-binding site in these target genes.

Next we inhibited the expression of TGF- $\beta$  and SMAD3 by using siRNAs that specifically target TGF- $\beta$  and SMAD3 (Fig. 4f). Down-regulation of TGF- $\beta$  and SMAD3 independently of miR-LAT also confers resistance to apoptosis (Fig. 4f), whereas a control scrambled siRNA did not show any such effect (Fig. 4f). This suggests that inhibition of the TGF- $\beta$  pathway by miR-LAT is sufficient for its anti-apoptotic effect.

Furthermore, we analysed the effect of miR-LAT on TGF- $\beta$ /SMAD-dependent transcription using SBE4-Luc, a TGF- $\beta$ -responsive reporter gene construct that contains four copies of the SMAD-binding element (SBE)<sup>15</sup>. As shown in Fig. 4g, pcDNA-PstMlu inhibited TGF- $\beta$ -mediated induction of the SBE4-Luc construct. The LAT deletion construct ( $\Delta$ Sty) had no effect on induction of the SBE4-Luc reporter gene as compared to pcDNA (Fig. 4g). LAT expression had no effect on TGF- $\beta$ -mediated induction of the SBE4-Luc reporter gene in the presence of 2'-O-methyl antisense oligonucleotides specific for miR-LAT (Fig. 4h). To elucidate further the role of miR-LAT in TGF- $\beta$  signalling, SY5Y cells were co-transfected with a synthetic double-stranded RNA oligonucleotide corresponding to mature miR-LAT along with the SBE4-Luc reporter gene. As shown in Fig. 4i, the synthetic miR-LAT oligonucleotide caused a significant reduction in TGF- $\beta$ -mediated induction of SBE4-Luc reporter gene activity as compared to a control (scrambled) double-stranded RNA oligonucleotide.

Taken together, our results show that the HSV-1 *LAT* gene (position +415 to +475) codes for a miRNA that can inhibit apoptosis. The *LAT* gene maintains latent infections by protecting infected neurons from undergoing apoptosis<sup>4</sup>. Furthermore, it has been shown that an unrelated anti-apoptotic gene can substitute for the HSV-1 *LAT* gene during reactivation of the virus in mice<sup>16,17</sup>. Our results suggest that the anti-apoptotic miRNA miR-LAT, encoded by the *LAT* region of HSV-1, has an important role in the survival of latently infected neurons and thus contributes to the persistence of



**Figure 4 | Modulation of TGF- $\beta$  signalling by miR-LAT.** **a**, The predicted duplex of miR-LAT and its target site in the 3' UTR of TGF- $\beta$  (top panel) or the 3' UTR of SMAD3 (bottom panel); canonical base pairs are marked with red circles; mismatches are shown as black circles. The sequence alignments in the middle panel show nucleotide conservation of the miR-LAT-binding site within the 3' UTR of TGF- $\beta$ . **b**, Top panel: RT-PCR analysis of the transcript levels of TGF- $\beta$ , SMAD3, TGF- $\beta$  receptor I and GAPDH. Bottom panel: western blot analysis for SMAD3, TGF- $\beta$  receptor I and  $\beta$ -actin. **c**, Top panel: SY5Y cells were co-transfected with *Renilla* luciferase constructs containing the full-length 3' UTRs of the indicated genes along with pSV $\beta$ -gal ( $\beta$ -gal), pcDNA or PstMlu. Data represent the mean of normalized *Renilla* luciferase/ $\beta$ -gal activities of four independent experiments  $\pm$  s.d. RLU, relative luciferase units. Lower panel: transcript levels of *Renilla* luciferase and firefly luciferase (as a normalization control) mRNAs as determined by RT-PCR. **d**, SY5Y cells were transfected with the indicated UTR constructs. Data represent the mean of normalized *Renilla* luciferase/ $\beta$ -gal activities of four independent experiments  $\pm$  s.d. **e**, Top panel: SY5Y cells were transfected with *Renilla* luciferase constructs containing the indicated wild-type or mutated target sites along with pcDNA or PstMlu. Data represent the mean of normalized *Renilla* luciferase/ $\beta$ -gal activities of four independent experiments  $\pm$  s.d. Bottom panel: transcript levels of *Renilla* luciferase and firefly luciferase (as a normalization control) as determined by RT-PCR. **f**, Top panel: SY5Y cells were treated with cisplatin, 24 h after transfection, and the percentage of apoptosis was determined. Data represent the mean of four independent experiments  $\pm$  s.d. Bottom panel: RT-PCR analysis of the transcript levels of TGF- $\beta$ , SMAD3 and GAPDH. **g-i**, SY5Y cells were transfected as indicated along with the SBE4-Luc reporter gene. Cells were treated with or without recombinant human TGF- $\beta$  and analysed for luciferase activity. Data represent the mean of four independent experiments  $\pm$  s.d.

infection. Recently it has been shown that mammalian cells can use RNA interference (RNAi) mechanisms to restrict viral propagation<sup>18,19</sup>. Some viruses have, in turn, evolved mechanisms to counter these RNA-mediated anti-viral responses of the host cell<sup>18,19</sup>. Our results suggest that miR-LAT inhibits apoptosis by modulation of TGF- $\beta$  signalling. Thus, HSV uses the RNAi pathway to regulate host cell apoptosis to ensure survival of infected neurons and maintenance of latent infection. Such a mechanism would circumvent the need for the expression of a viral protein during the latent infection and thus help the virus to evade immune detection. Our results highlight a novel role for the RNAi pathway in host–virus interactions.

## METHODS

See Supplementary Methods for information regarding detailed protocols.

**Cells and virus.** SY5Y cells were maintained in RPMI medium with 10% fetal calf serum and antibiotics. HeLa cells were grown in DMEM medium supplemented with 5% fetal calf serum and antibiotics. Plasmids having the PstMlu fragment of LAT and its deletion mutant  $\Delta$ Sty have been described previously<sup>5</sup>. The wild-type strain of HSV-1 (17+) and the mutant virus ( $\Delta$ Sty) have been described earlier<sup>5</sup>.

**Apoptosis assay.** At 24 h after transfection cells were treated with cisplatin (10  $\mu$ M) or etoposide (1–100  $\mu$ M), followed by staining with Annexin V FITC (BD Biosciences). Cells showing FITC staining, loss of cell volume, loss of refractility and membrane blebbing were scored as apoptotic. At least 500 cells were counted in each well.

**RNA isolation, RT–PCR, northern blotting and western analysis.** RNA was isolated using Trizol reagent (Invitrogen). RNA was reverse transcribed using Superscript II RT enzyme (Invitrogen). Northern blot analysis of 50–60  $\mu$ g total RNA was done as described previously<sup>21</sup> (see also [http://web.wi.mit.edu/bartel/pub/protocols/miRNA\\_Nrtrms\\_Protocol.pdf](http://web.wi.mit.edu/bartel/pub/protocols/miRNA_Nrtrms_Protocol.pdf)). Western blot analysis was performed on whole-cell lysates using Dicer polyclonal antibody<sup>20</sup> and  $\beta$ -actin antibody (Amersham). SMAD3 and TGF- $\beta$  receptor I antibodies were from Abcam.

**Reporter assays.** Cells were co-transfected with pcDNA or PstMlu (1  $\mu$ g) and the respective reporter constructs (500 ng) along with pSV- $\beta$ gal (100 ng). For induction of TGF- $\beta$  signalling, 24 h after transfection the cells were treated with 10 ng ml<sup>-1</sup> recombinant TGF- $\beta$ 1 protein (R&D systems). Reporter activity was measured using appropriate assay kits (Promega).

**miRNA cloning.** Small RNA molecules were cloned from PstMlu-transfected cells according to the protocols described previously<sup>22</sup>.

**Computational methods.** RNAMOT<sup>8</sup> was used to search for RNA hairpin. The detected duplexes were further analysed by MFOLD<sup>9</sup>. Targets of miR-LAT were identified using the software miRanda<sup>11</sup>. Statistical analysis was performed using JMPIN 4 software. One-way analysis of variance (ANOVA) was performed at an alpha level of 0.01–0.05 to compare all the variants in a data set.

Received 27 March; accepted 25 April 2006.

Published online 31 May 2006.

1. Ambros, V. MicroRNAs: tiny regulators with great potential. *Cell* **107**, 823–826 (2001).
2. Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism and function. *Cell* **116**, 281–297 (2004).
3. Jones, C. Herpes simplex virus type 1 and bovine herpesvirus 1 latency. *Clin. Microbiol. Rev.* **16**, 79–95 (2003).
4. Perng, G. C. *et al.* Virus-induced neuronal apoptosis blocked by the herpes simplex virus latency-associated transcript. *Science* **287**, 1500–1503 (2000).
5. Ahmed, M., Lock, M., Miller, C. G. & Fraser, N. W. Regions of the herpes

simplex virus type 1 latency-associated transcript that protect cells from apoptosis *in vitro* and protect neuronal cells *in vivo*. *J. Virol.* **76**, 717–729 (2002).

6. Inman, M. *et al.* Region of herpes simplex virus type 1 latency-associated transcript sufficient for wild-type spontaneous reactivation promotes cell survival in tissue culture. *J. Virol.* **75**, 3636–3646 (2001).
7. Bannasser, Y., Le, S. Y., Benkirane, M. & Jeang, K. T. Evidence that HIV-1 encodes an siRNA and a suppressor of RNA silencing. *Immunity* **22**, 607–619 (2005).
8. Laferriere, A., Gautheret, D. & Cedergren, R. An RNA pattern matching program with enhanced performance and portability. *Comput. Appl. Biosci.* **10**, 211–212 (1994).
9. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).
10. Tomari, Y. & Zamore, P. D. Perspective: machines for RNAi. *Genes Dev.* **19**, 517–529 (2005).
11. John, B. *et al.* Human MicroRNA targets. *PLoS Biol.* **2**, e363 (2004).
12. Derynck, R., Akhurst, R. J. & Balmain, A. TGF- $\beta$  signaling in tumor suppression and cancer progression. *Nature Genet.* **29**, 117–129 (2001).
13. Feng, X. H. & Derynck, R. Specificity and versatility in TGF- $\beta$  signaling through Smads. *Annu. Rev. Cell Dev. Biol.* **21**, 659–693 (2005).
14. Schuster, N. & Kriegelstein, K. Mechanisms of TGF- $\beta$ -mediated apoptosis. *Cell Tissue Res.* **307**, 1–14 (2002).
15. Zawel, L. *et al.* Human Smad3 and Smad4 are sequence-specific transcription activators. *Mol. Cell* **1**, 611–617 (1998).
16. Jin, L. *et al.* Identification of herpes simplex virus type 1 latency-associated transcript sequences that both inhibit apoptosis and enhance the spontaneous reactivation phenotype. *J. Virol.* **77**, 6556–6561 (2003).
17. Jin, L. *et al.* A herpes simplex virus type 1 mutant expressing a baculovirus inhibitor of apoptosis gene in place of latency-associated transcript has a wild-type reactivation phenotype in the mouse. *J. Virol.* **79**, 12286–12295 (2005).
18. Sullivan, C. S. & Ganem, D. MicroRNAs and viral infection. *Mol. Cell* **20**, 3–7 (2005).
19. Schutz, S. & Sarnow, P. Interaction of viruses with the mammalian RNA interference pathway. *Virology* **344**, 151–157 (2006).
20. Chendrimada, T. P. *et al.* TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature* **436**, 740–744 (2005).
21. Lau, N. C., Lim, L. P., Weinstein, E. G. & Bartel, D. P. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* **294**, 797–799 (2001).
22. Pfeffer, S., Lagos-Quintana, M. & Tuschl, T. in *Current Protocols in Molecular Biology* (eds Ausbel, F. M. *et al.*) 26.4.1–26.4.18 (Wiley Interscience, New York, 2003).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank K.-T. Jeang for the GFP-shRNA construct; Z. Mourelatos for the PRL-TK vector; and R. Sheikhatrar for the Dicer polyclonal antibody. We thank B. Brunk for help with the computational analysis. We also thank T. Chendrimada for advice and help with miRNA protocols. We acknowledge S. Berges for critical review of the manuscript. This work was supported by a National Institute of Health grant. A.G.H. is supported by an NSF Career Award Grant.

**Author Contributions** A.G. conceived the project and carried out all experiments described. J.J.G. provided technical assistance. P.S. and A.G.H. developed and applied computational algorithm for miRNA detection and target prediction. N.W.F. directed and supervised the experimental work and interpretation of data. The manuscript was prepared by A.G. and N.W.F.

**Author Information** Reprints and permissions information is available at [npg.nature.com/reprintsandpermissions](http://npg.nature.com/reprintsandpermissions). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to N.W.F. ([nwfraser@mail.med.upenn.edu](mailto:nwfraser@mail.med.upenn.edu)).

## LETTERS

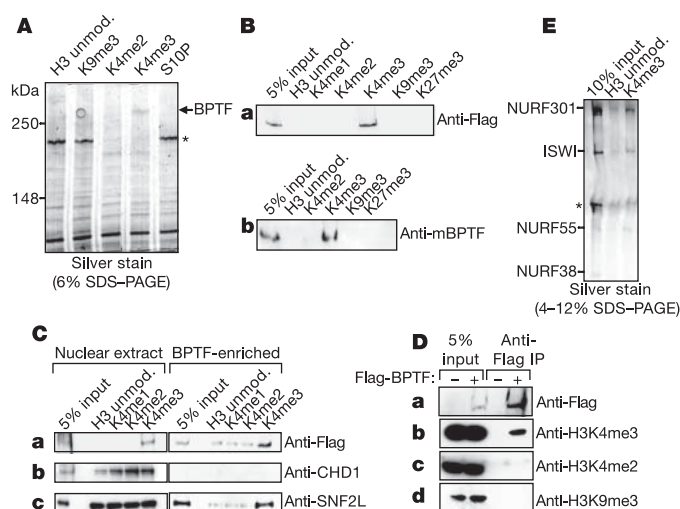
# A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling

Joanna Wysocka<sup>1</sup>, Tomek Swigut<sup>2</sup>, Hua Xiao<sup>4</sup>, Thomas A. Milne<sup>1</sup>, So Yeon Kwon<sup>5</sup>, Joe Landry<sup>4</sup>, Monika Kauer<sup>1</sup>, Alan J. Tackett<sup>3</sup>, Brian T. Chait<sup>3</sup>, Paul Badenhurst<sup>5</sup>, Carl Wu<sup>4</sup> & C. David Allis<sup>1</sup>

Lysine methylation of histones is recognized as an important component of an epigenetic indexing system demarcating transcriptionally active and inactive chromatin domains. Trimethylation of histone H3 lysine 4 (H3K4me3) marks transcription start sites of virtually all active genes<sup>1–4</sup>. Recently, we reported that the WD40-repeat protein WDR5 is important for global levels of H3K4me3 and control of *HOX* gene expression<sup>5</sup>. Here we show that a plant homeodomain (PHD) finger of nucleosome remodelling factor (NURF), an ISWI-containing ATP-dependent chromatin-remodelling complex, mediates a direct preferential association with H3K4me3 tails. Depletion of H3K4me3 causes partial release of the NURF subunit, BPTF (bromodomain and PHD finger transcription factor), from chromatin and defective recruitment of the associated ATPase, SNF2L (also known as ISWI and SMARCA1), to the *HOXC8* promoter. Loss of BPTF in *Xenopus* embryos mimics WDR5 loss-of-function phenotypes, and compromises spatial control of *Hox* gene expression. These results strongly suggest that WDR5 and NURF function in a common biological pathway *in vivo*, and that NURF-mediated ATP-dependent chromatin remodelling is directly coupled to H3K4 trimethylation to maintain *Hox* gene expression patterns during development. We also identify a previously unknown function for the PHD finger as a highly specialized methyl-lysine-binding domain.

Recent genomic-scale analyses of histone modifications indicate that H3K4 trimethylation (H3K4me3) is preferentially associated with the transcription start sites of active genes<sup>2–4</sup>. Several factors have been proposed to recognize H3K4 methylated tails and mediate transcriptional activation (for example, CHD1 (refs 6, 7), Iswl (ref. 8) and WDR5 (ref. 5)). However, these factors do not distinguish between di- and tri-methylated H3K4. In a previous study, we postulated that an effector recognizing H3K4me3 should act downstream of WDR5 (ref. 5).

To identify proteins that preferentially bind H3K4me3, nuclear extracts from HEK293 cells were incubated with unmodified and modified histone H3 peptides. This analysis revealed that a polypeptide >300 kDa specifically bound to the H3K4me3 peptide (Fig. 1A). This protein was identified by mass spectrometry to be BPTF. BPTF is the largest subunit of NURF<sup>9,10</sup>, a chromatin-remodelling complex closely involved in transcription<sup>9–17</sup> (see Supplementary Fig. S1). To confirm the specificity of the above interaction, nuclear extracts from a cell line expressing Flag-tagged human BPTF or from mouse NIH3T3 cells were used in peptide pull-down assays with a series of H3 peptides. As shown in Fig. 1B, human (Fig. 1B, panel a) and mouse (Fig. 1B, panel b) BPTF specifically associates with H3K4me3 peptide, but not H3K4me2 peptide or H3 peptides trimethylated at other residues (for example, K9 and K27).



**Figure 1 | Human and *Drosophila* NURF specifically associate with histone H3 trimethylated at K4.** **A**, Peptide pull-down assays were performed using HEK293 cell nuclear extracts and unmodified (unmod.), methylated or phosphorylated H3 peptides as indicated and visualized by silver staining. A polypeptide specifically enriched in the H3K4me3 pull down was analysed by mass spectrometry and identified as BPTF (arrow). The polypeptide indicated with an asterisk corresponds to Mi-2, a subunit of the NuRD complex previously reported to associate with unmodified, but not K4 methylated, H3 tail<sup>27,28</sup>. S10P, H3 peptide phosphorylated at serine 10. **B**, Human and mouse BPTF specifically associate with H3K4me3 peptide. Nuclear extracts from HEK293 cells stably expressing Flag-tagged human BPTF (**a**), or from NIH3T3 cells (**b**), were used in peptide pull-down assays and analysed by immunoblotting with anti-Flag (**a**) or anti-mouse BPTF (mBPTF; **b**) antibodies. **C**, BPTF targets SNF2L to H3K4me3 in a CHD1-independent manner. Pull-down assays were performed using Flag-BPTF HEK293 nuclear extract or the same enriched for BPTF by Flag immunoprecipitation. Results were analysed by immunoblotting with the indicated antibodies (**a–c**). **D**, BPTF associates with H3K4me3 nucleosomes. Mononucleosomal fractions were purified from mock-transfected or Flag-BPTF HEK293 cell lines and used for immunoprecipitation (IP) with anti-Flag antibody; immunoprecipitates were analysed by immunoblotting with the indicated antibodies (**a–d**). **E**, *Drosophila* NURF recognizes H3K4me3. Reconstituted, gradient-purified *Drosophila* NURF was used in a histone-tail peptide pull-down assay. Input and bound proteins were resolved by SDS-PAGE and visualized by silver staining. NURF subunits are indicated on the left. The asterisk indicates a NURF301 degradation product, which is not enriched in the H3K4me3 pull down.

<sup>1</sup>Laboratory of Chromatin Biology, <sup>2</sup>Laboratory of Molecular Vertebrate Embryology, and <sup>3</sup>Laboratory of Mass Spectrometry and Gaseous Ion Chemistry, The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA. <sup>4</sup>Laboratory of Molecular Cell Biology, National Cancer Institute, NIH, Bethesda, Maryland 20814, USA. <sup>5</sup>Institute of Biomedical Research, University of Birmingham, Edgbaston B15 2TT, UK.



Next, we verified that this association is independent of CHD1 (ref. 7), and also targets the SNF2L/ISWI ATPase to the H3K4 trimethylated tail. Immunoblot analysis indicated that BPTF preferentially associated with H3K4me3 in both nuclear extracts and BPTF-enriched material (Fig. 1C, panel a). In contrast, CHD1 bound to all forms of methylated H3K4 (with a preference for di- and tri-methyl), but was not detected in the BPTF-enriched input or peptide elutions (Fig. 1C, panel b). SNF2L/ISWI associated with the H3 tail in a methylation-independent manner (Fig. 1C, panel c), but in BPTF-enriched material SNF2L/ISWI was specifically enriched in the H3K4me3 pull down. These results indicate that (1) BPTF associates with H3K4me3 independently of CHD1, (2) BPTF can target SNF2L/ISWI to H3K4me3, and (3) the majority of SNF2L/ISWI is not associated with BPTF and does not bind preferentially to H3K4me3. Furthermore, tail peptide pull-down assays using recombinant *Drosophila* NURF complex indicates that the association between NURF and H3K4me3 is conserved (Fig. 1E).

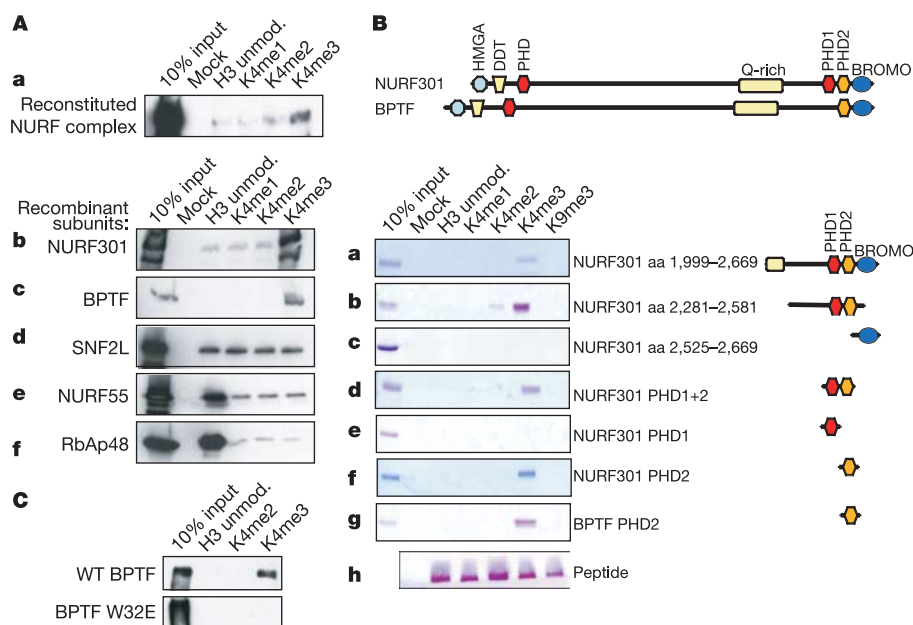
To confirm that these *in vitro* interactions occur in the context of chromatin *in vivo*, we tested association of BPTF with H3K4 methylated nucleosomes. Chromatin was isolated from either Flag-BPTF-expressing (+) or non-expressing (–) HEK293 cells and solubilized with micrococcal nuclease. Mononucleosome fractions were pooled and immunoprecipitated with anti-Flag antibody, followed by Flag peptide elution and immunoblotting with antibodies against (1) Flag (Fig. 1D, panel a), (2) H3K4me3 (Fig. 1D, panel b), (3) H3K4me2 (Fig. 1D, panel c) or (4) H3K9me3 (Fig. 1D, panel d). BPTF specifically co-immunoprecipitated with the H3K4 trimethylated mononucleosomes (compare panels b–d of Fig. 1D), suggesting that NURF binds H3K4me3 in a nucleosomal context. Antibodies suitable for *in situ* analyses for NURF301 (the *Drosophila* orthologue of BPTF; also known as E(bx) (enhancer of bithorax)) do not exist, but double immunostaining of salivary gland polytene

chromosomes showed some overlap between sites of ISWI binding and sites of H3K4me3 (data not shown), consistent with the possibility that the NURF complex associates with H3K4me3.

We next verified that the large subunit of NURF dictates H3K4me3-binding by testing the individual binding specificities of BPTF, NURF301 or other subunits of the *Drosophila* and human NURF complexes. Haemagglutinin (HA)- or Flag-tagged components of NURF were individually expressed in SF9 cells, and extracts used in histone tail peptide pull-down assays. NURF301 and BPTF were enriched in the H3K4me3 pull down, similar to the gradient-purified, reconstituted NURF complex (Fig. 2A, panels a–c). Human SNF2L/ISWI bound irrespective of the peptide H3K4 methylation status (Fig. 2A, panel d), whereas *Drosophila* NURF55 and its human orthologue RbAp48 (also known as RBBP4) preferentially bound to the unmodified H3 peptide (Fig. 2A, panels e and f). Recombinant NURF38 did not bind any of the H3 peptides tested (data not shown).

BPTF and NURF301 are large, multidomain proteins. To identify the region responsible for H3K4me3-binding, glutathione S-transferase (GST) fusions that span NURF301 (ref. 15) were expressed in *Escherichia coli*, purified and tested in peptide pull-down assays. Only the most carboxy-terminal region (residues 1,999–2,669; see schematics in Fig. 2B), specifically bound to H3K4me3 (Fig. 2B, panel a; data not shown). Further deletion narrowed the region of interaction to the second (bromodomain-proximal) PHD finger (PHD2), which is conserved in human BPTF (Fig. 2B, panels e–g).

We next confirmed that the PHD finger is obligatory—in the context of the full-length protein—for H3K4me3-binding. Cells were transfected with either Flag-tagged wild-type (WT) human BPTF, or Flag-tagged BPTF containing a W32E substitution in the PHD finger (PHD-W32E; see ref. 18), and extracts were subjected to a peptide pull-down assay. The W32E substitution abolished recognition of



**Figure 2 | The bromodomain-proximal PHD finger of BPTF/NURF301 binds to H3 K4me3.** **A**, BPTF/NURF301 confers H3K4me3 recognition. Reconstituted, gradient-purified *Drosophila* NURF (**a**), extracts from SF9 cells expressing individual tagged components of *Drosophila* NURF (**b**, **e**), or human NURF components (**c**, **d**, **f**), were used in pull-down assays followed by immunoblotting. **B**, The PHD finger of BPTF/NURF301 is sufficient for association with H3K4me3. Top, schematics show *Drosophila* NURF301 and human BPTF. HMGA, high-mobility group box A domain; DDT, DNA-binding homeobox and different transcription factors domain; Q-rich, glutamine-rich domain; BROMO, bromodomain. Bottom, peptide

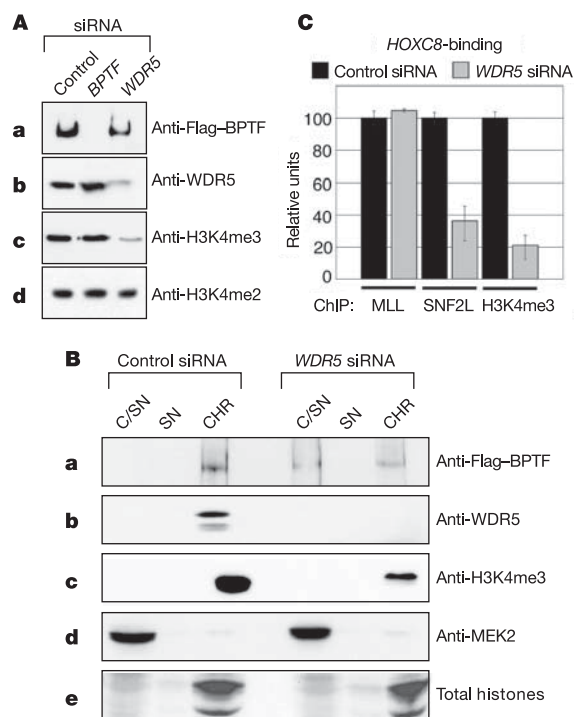
pull-down assays using recombinant GST–NURF301 (**a**–**f**) and GST–BPTF (**g**) fusion proteins containing different regions of the NURF301/BPTF C terminus. H3 peptides used in the assays are indicated. Bound proteins were analysed by Coomassie-blue staining. Equivalent peptide loading was ensured by silver staining following SDS–PAGE (**h**). **C**, The PHD finger of BPTF is necessary for H3K4me3-binding. Nuclear extracts from cells transfected with a construct encoding Flag-tagged wild-type (WT) human BPTF, or BPTF containing a W32E substitution in the PHD finger, were used in pull-down assays with unmodified and H3K4me3 peptides. Bound proteins were probed with an anti-Flag antibody.

H3K4me3 by BPTF (Fig. 2C). Taken together, these results indicate that the bromodomain-proximal PHD finger is both necessary and sufficient for H3K4me3 association.

Although NURF is known to be directly recruited by site-specific transcription factors (for example, in *Drosophila*, GAGA factor<sup>15</sup> and the ecdysone receptor<sup>17</sup>), its association with chromatin is likely to be stabilized by H3K4me3 (see Supplementary Fig. S3A). This model predicts that (1) loss of BPTF, unlike loss of WDR5, should not affect H3K4me3, (2) loss of H3K4me3 should reduce chromatin association of BPTF, and (3) WDR5 and BPTF function in a common biological pathway.

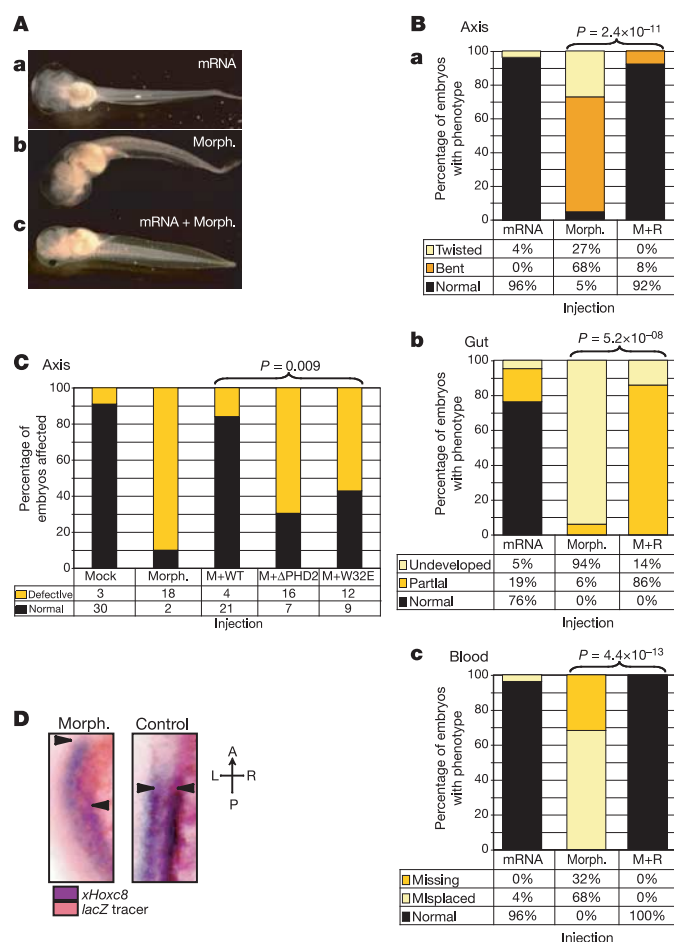
To determine whether BPTF directly controls H3K4 methylation, HEK293 cells transfected with BPTF short interfering RNA (siRNA) were compared with cells transfected with non-coding and WDR5 siRNA controls. BPTF siRNA transfection caused efficient knockdown of BPTF (Fig. 3A, panel a), but did not affect either WDR5 levels (Fig. 3A, panel b) or global levels of H3K4me (Fig. 3A, panels c and d).

To test the second prediction, HEK293 cells stably expressing Flag-BPTF were transfected with either control or WDR5 siRNA, and then fractionated to separate subcellular fractions<sup>19</sup>. WDR5 siRNA treatment ablated WDR5 from all fractions and reduced H3K4me3 levels (Fig. 3B, panels b and c). WDR5 knockdown caused the partial loss of BPTF from chromatin (~40–50% loss), and the appearance of BPTF in the soluble cytosolic/nuclear fraction. Partial loss of BPTF–chromatin association may be due to incomplete loss of H3K4me3 and/or redundancy in other modes of chromatin association. Nevertheless,



**Figure 3 | WDR5 regulates BPTF chromatin association.** **A**, BPTF knockdown does not affect H3K4 methylation. HEK293 Flag-BPTF cells were transfected with control siRNAs, a BPTF siRNA pool, or a WDR5 siRNA pool and probed with the indicated antibodies (**a–d**). **B**, WDR5 knockdown results in partial release of BPTF from chromatin. Biochemical fractionation<sup>19</sup> was used to generate three fractions: C/SN (cytosolic proteins and some soluble nuclear components); SN (remaining soluble nuclear components); and, CHR (chromatin). Fractions were probed with the antibodies indicated (**a–d**) or stained with Coomassie blue for analysis of total histones (**e**). **C**, Association of SNF2L with the *HOXC8* locus requires WDR5. ChIP assays with anti-MLL, anti-SNF2L and anti-H3K4me3 antibodies were performed using HEK293 cells treated with control (black bars) and WDR5 (grey bars) siRNAs. Signals were quantified relative to inputs (see Methods). Error bars indicate s.d.

our results suggest that H3K4me3 stabilizes association of BPTF/NURF with chromatin (see the model in Supplementary Fig. S3B). This conclusion is additionally supported by chromatin immunoprecipitation (ChIP) analysis of the *HOXC8* locus in control versus



**Figure 4 | Loss of BPTF phenocopies loss of WDR5 in axial, gut and blood defects during *Xenopus* development.** **A**, Feeding-stage tadpoles, derived from two-cell-stage embryos that were injected in one blastomere with *xBPTF* morpholino (**b**), *hBPTF* mRNA (**a**), or *xBPTF* morpholino and *hBPTF* mRNA (**c**). Tadpoles derived from the morpholino-injected embryos show axial, blood and gut patterning defects (see the text) that are rescued by co-injection of *hBPTF* mRNA. **B**, Penetrance of phenotypes shown in **A**. The occurrence of axial deformities (**a**), gut patterning defects (**b**), and blood defects (**c**) were scored in morpholino-injected embryos (“Morph.”), embryos co-injected with morpholino and *hBPTF* mRNA (“M+R”), or those injected with *hBPTF* mRNA alone (“mRNA”); *P*-values are indicated above each plot. **C**, Quantitation of rescue by WT BPTF or forms of BPTF defective in H3K4me3-binding. Axial deformities were scored in morpholino-injected embryos (“Morph.”), and embryos co-injected with morpholino and mRNAs encoding either WT *hBPTF* (“M+WT”), *hBPTF* with C-terminal truncation (“M+ΔPHD2”), or *hBPTF* containing a point mutation in the PHD domain that abolishes H3K4me3 recognition (“M+W32E”). Amount of injected RNA was standardized between samples. Results were plotted as proportions in the total population counted; *P*-values are indicated above each plot, and the absolute number of embryos in each category are shown below. Similar analyses were done for blood and gut defects (Supplementary Fig. S4). **D**, Whole-mount *in situ* detection of the *xHoxc8* transcript. Dorsal view of the mid-trunk region of stage-30 tadpoles derived from embryos injected with *xBPTF* morpholino (left panel) or injected with a control (right panel). Purple stain indicates *xHoxc8* expression, which is most prominent in the neural tube, and red stain indicates the lineage tracer (*lacZ*). Arrows indicate misalignment of the *xHoxc8* expression domains between the left and right side of the neural tube in morpholino-injected, but not control, embryos. Images are shown in the same orientation. A–P, anterior–posterior axis; L–R, left–right axis.

WDR5 siRNA-treated cells. As ChIP-grade BPTF antibodies are unavailable, we used antibodies against SNF2L/ISWI and determined that SNF2L/ISWI, but not MLL (mixed-lineage leukemia), binding to the *HOXC8* locus, was dependent on the presence of WDR5 and correlated with trimethylation of the locus at H3K4 (Fig. 3C). Furthermore, we demonstrated that binding of SNF2L/ISWI to the *Hoxc8* locus is reduced in *MLL*<sup>-/-</sup>, compared with *MLL*<sup>+/+</sup>, mouse embryonic fibroblasts (Supplementary Fig. S2C). Consistent with previous data<sup>20</sup>, H3K4 methylation of the *Hoxc8* locus is dependent on MLL (Supplementary Fig. S2B). The correlation between binding of SNF2L/ISWI and H3K4me3 at *HOXC8* supports the hypothesis that H3K4me3 stabilizes BPTF–SNF2L complexes at target genes.

We next investigated whether loss of BPTF in *Xenopus* embryos phenocopies defects previously described for loss of WDR5 (ref. 5). One blastomere of the two-cell-stage *Xenopus laevis* embryo was injected with either a morpholino oligonucleotide targeting *Xenopus BPTF* (*xBPTF* morpholino), *xBPTF* morpholino and human *BPTF* messenger RNA (*hBPTF* mRNA), or *hBPTF* mRNA alone. Injected embryos were allowed to develop to the feeding tadpole stages. *xBPTF* morpholino-injected tadpoles showed axial defects on the injected side of the body resulting in a distinctive bend in the embryo (Fig. 4A, panel b), as well as gut and blood defects (see Fig. 4B, panels a–c, for penetrance analysis). All phenotypes were highly penetrant and were rescued by the co-injection of *hBPTF* mRNA along with the *xBPTF* morpholino (Fig. 4A, panel c; Fig. 4B). Injection of the *hBPTF* mRNA alone had no significant effect (Fig. 4A, panel a; Fig. 4B), and phenotype penetrance and severity was dose-dependent (data not shown). The mRNA-mediated rescue of these defects indicates that the phenotypes are due to BPTF protein knockdown and not to non-specific effects.

Notably, the phenotypes observed in *xBPTF* morpholino-injected tadpoles are strikingly similar to those resulting from a *xWDR5* morpholino<sup>5</sup>. Like *xWDR5* knockdown<sup>5</sup>, *xBPTF* knockdown causes deregulation of *xHoxc8* expression. Two-cell-stage embryos were injected into one blastomere with *xBPTF* morpholino and *lacZ* mRNA as a lineage tracer. *xBPTF* knockdown resulted in the posteriorization of *xHoxc8* expression by several somite lengths, indicating that BPTF, like WDR5, is important for control of *Hox* gene expression (Fig. 4D). In support of this, *Drosophila* NURF301 (a Trithorax group member) is required for homeotic gene expression<sup>16</sup>.

To address whether recognition of H3K4me3 is important for BPTF function *in vivo*, rescue experiments were performed in parallel using an mRNA encoding (1) WT hBPTF, (2) hBPTF with a C-terminal truncation that deletes the second (bromodomain-proximal) PHD finger along with the bromodomain ( $\Delta$ PHD2), or (3) hBPTF containing a single point mutation resulting in loss of H3K4me3 recognition (PHD-W32E; see Fig. 2C). The  $\Delta$ PHD2 truncation and the PHD-W32E substitution decreased rescue of axial (Fig. 4C), blood and gut (Supplementary Fig. S4) phenotypes, as compared with WT mRNA. These results demonstrate that recognition of H3K4me3 is essential for NURF function *in vivo*.

Our work demonstrates that two conserved chromatin regulators—WDR5, an adaptor protein essential for H3K4 trimethylation, and NURF, an archetypal remodelling complex—function in a common biological pathway. As H3K4 trimethylation correlates with transcription start sites, we favour the view that NURF recruitment to promoters, mediated by sequence-specific transcription factors<sup>9,17</sup> and stabilized by binding to H3 tails trimethylated at K4, modulates transcription initiation through ATP-dependent nucleosome sliding.

Our study identifies the PHD finger as a trimethyl-lysine-binding domain, capable of highly specialized interpretation of the H3K4 methylation mark. The methyl-lysine association is likely to be a general function of the subset of PHD fingers containing the aromatic cluster (see ref. 18). In support of this, the PHD finger of the tumour suppressor protein ING2 also recognizes methylated H3K4 (see ref. 21). PHD fingers are present in multiple chromatin-associated

proteins<sup>22</sup>, often in conjunction with a bromodomain—a module that recognizes acetylated histone tails<sup>23</sup>. Moreover, PHD fingers and bromodomains have been shown to functionally cooperate<sup>24,25</sup>. Notably, the H3K4me3-binding PHD finger of BPTF/NURF301 is situated in close proximity to its extreme C-terminal bromodomain, suggesting the existence of a double-domain module, which would most likely recognize a trimethyl (K4)-acetyl pattern of modifications. The higher efficiency of rescue with the PHD-W32E point mutant, compared with the C-terminal truncation, suggests that the PHD finger and the bromodomain may cooperate in mediating BPTF function.

## METHODS

**Pull-down assays.** Pull-down assays with extracts and recombinant proteins were performed as described previously<sup>5</sup>. Nuclear extracts were prepared from HEK293 cells or NIH3T3 cells using the Dignam protocol<sup>26</sup>. Briefly, 10<sup>8</sup> cells were used per pull-down assay. Salt and Triton-X100 concentrations were 150 mM and 0.2% (v/v), respectively. Nuclear extracts were precleared with avidin beads and incubated with peptide prebound to avidin beads for 3 h at 4 °C. Approximately 5  $\mu$ g of peptide was used per pull down. Beads were washed eight times with 20 mM HEPES, pH 7.9, 250 mM KCl, 0.2% (v/v) Triton-X100, 1 mM phenylmethylsulphonyl fluoride (PMSF), and containing protease inhibitor cocktail (Roche). The final wash was performed with 4 mM HEPES, pH 7.9, 10 mM NaCl, 1 mM PMSF, and containing protease inhibitor cocktail. Bound proteins were eluted from the resin twice using 100 mM glycine, pH 2.8. Eluates were combined, neutralized using 0.1 $\times$  volume of 1 M Tris-HCl, pH 8, and analysed by SDS-PAGE. The HEK293 Flag-BPTF cell line was prepared by co-transfection of pCMV-Flag-BPTF with a 20-fold excess of PAC vector, followed by puromycin selection at 4 mg ml<sup>-1</sup> and subcloning of individual colonies. All histone peptides (amino acids 1–20) were synthesized by the Rockefeller University Proteomics Resource Center.

Details of additional methods used in this study are included in Supplementary Information.

Received 24 February; accepted 17 April 2006.

Published online 21 May 2006.

- Santos-Rosa, H. *et al.* Active genes are tri-methylated at K4 of histone H3. *Nature* **419**, 407–411 (2002).
- Schneider, R. *et al.* Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. *Nature Cell Biol.* **6**, 73–77 (2004).
- Bernstein, B. E. *et al.* Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120**, 169–181 (2005).
- Pokholok, D. K. *et al.* Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* **122**, 517–527 (2005).
- Wysocka, J. *et al.* WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell* **121**, 859–872 (2005).
- Pray-Grant, M. G., Daniel, J. A., Schieltz, D., Yates, J. R. & Grant, P. A. Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation. *Nature* **433**, 434–438 (2005).
- Sims, R. J. *et al.* Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains. *J. Biol. Chem.* **280**, 41789–41792 (2005).
- Santos-Rosa, H. *et al.* Methylation of histone H3 K4 mediates association of the Isw1p ATPase with chromatin. *Mol. Cell* **12**, 1325–1332 (2003).
- Tsukiyama, T. & Wu, C. Purification and properties of an ATP-dependent nucleosome remodeling factor. *Cell* **83**, 1011–1020 (1995).
- Barak, O. *et al.* Isolation of human NURF: a regulator of *Engrailed* gene expression. *EMBO J.* **22**, 6089–6100 (2003).
- Tsukiyama, T., Becker, P. B. & Wu, C. ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor. *Nature* **367**, 525–532 (1994).
- Mizuguchi, G., Tsukiyama, T., Wisniewski, J. & Wu, C. Role of nucleosome remodeling factor NURF in transcriptional activation of chromatin. *Mol. Cell* **1**, 141–150 (1997).
- Hamiche, A., Sandaltzopoulos, R., Gdula, D. A. & Wu, C. ATP-dependent histone octamer sliding mediated by the chromatin remodeling complex NURF. *Cell* **97**, 833–842 (1999).
- Mizuguchi, G., Vassilev, A., Tsukiyama, T., Nakatani, Y. & Wu, C. ATP-dependent nucleosome remodeling and histone hyperacetylation synergistically facilitate transcription of chromatin. *J. Biol. Chem.* **276**, 14773–14783 (2001).
- Xiao, H. *et al.* Dual functions of largest NURF subunit NURF301 in nucleosome sliding and transcription factor interactions. *Mol. Cell* **8**, 531–543 (2001).
- Badenhorst, P., Voas, M., Rebay, I. & Wu, C. Biological functions of the ISWI chromatin remodeling complex NURF. *Genes Dev.* **16**, 3186–3198 (2002).
- Badenhorst, P. *et al.* The *Drosophila* nucleosome remodeling factor NURF is



- required for Ecdysteroid signaling and metamorphosis. *Genes Dev.* **19**, 2540–2545 (2005).
18. Li, H. *et al.* Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature* advance online publication, doi:10.1038/nature04802 (21 May 2006).
  19. Méndez, J. & Stillman, B. Chromatin association of human origin recognition complex, cdc6, and minichromosome maintenance proteins during the cell cycle: assembly of prereplication complexes in late mitosis. *Mol. Cell. Biol.* **20**, 8602–8612 (2000).
  20. Milne, T. A. *et al.* MLL targets SET domain methyltransferase activity to *Hox* gene promoters. *Mol. Cell* **10**, 1107–1117 (2002).
  21. Shi, X. *et al.* ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature* advance online publication, doi:10.1038/nature04835 (21 May 2006).
  22. Biehn, M. The PHD finger, a nuclear protein-interaction domain. *Trends Biochem. Sci.* **31**, 35–40 (2006).
  23. Dhalluin, C. *et al.* Structure and ligand of a histone acetyltransferase bromodomain. *Nature* **399**, 491–496 (1999).
  24. Ragvin, A. *et al.* Nucleosome binding by the bromodomain and PHD finger of the transcriptional cofactor p300. *J. Mol. Biol.* **337**, 773–788 (2004).
  25. Zhou, Y. & Grummt, I. The PHD finger/bromodomain of NoRC interacts with acetylated histone H4K16 and is sufficient for rDNA silencing. *Curr. Biol.* **15**, 1434–1438 (2005).
  26. Dignam, J. D., Lebovitz, R. M. & Roeder, R. G. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* **11**, 1475–1489 (1983).
  27. Zegerman, P., Canas, B., Pappin, D. & Kouzarides, T. Histone H3 lysine 4 methylation disrupts binding of nucleosome remodeling and deacetylase (NuRD) repressor complex. *J. Biol. Chem.* **277**, 11621–11624 (2002).
  28. Nishioka, K. *et al.* Set9, a novel histone H3 methyltransferase that facilitates transcription by precluding histone tail modifications required for heterochromatin formation. *Genes Dev.* **16**, 479–489 (2002).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank E. Bernstein, W. Herr and E. Duncan for critical reading of the manuscript, O. Gozani for communicating unpublished results, and D. Reinberg, W. Herr, R. Roeder and J. Tamkun for CHD1, WDR5, MLL and ISWI antibodies, respectively. T.S. thanks A. H. Brivanlou for support. J.W. is a D. Runyon CRF Fellow, T.A.M. is a Canadian Institutes of Health Research Fellow, and J.L. is an ACS Postdoctoral Research Fellow. C.W., J.L. and H.X. are supported by the US National Cancer Institute Intramural Research Program, and S.Y.K. and P.B. are supported by the BBSRC. C.D.A. acknowledges a MERIT Award from the NIH.

**Author Information** Reprints and permissions information is available at [npg.nature.com/reprintsandpermissions](http://npg.nature.com/reprintsandpermissions). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.D.A. ([alliscd@rockefeller.edu](mailto:alliscd@rockefeller.edu)), and requests for materials relating to *Drosophila* and mammalian NURF to C.W. ([carlwu@helix.nih.gov](mailto:carlwu@helix.nih.gov)).

# Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF

Haitao Li<sup>1</sup>, Serge Ilin<sup>1</sup>, Wooikoon Wang<sup>1</sup>, Elizabeth M. Duncan<sup>2</sup>, Joanna Wysocka<sup>2</sup>, C. David Allis<sup>2</sup> & Dinshaw J. Patel<sup>1</sup>

Mono-, di- and trimethylated states of particular histone lysine residues are selectively found in different regions of chromatin, thereby implying specialized biological functions for these marks ranging from heterochromatin formation to X-chromosome inactivation and transcriptional regulation<sup>1–3</sup>. A major challenge in chromatin biology has centred on efforts to define the connection between specific methylation states and distinct biological read-outs impacting on function<sup>4</sup>. For example, histone H3 trimethylated at lysine 4 (H3K4me3) is associated with transcription start sites of active genes<sup>5–7</sup>, but the molecular ‘effectors’ involved in specific recognition of H3K4me3 tails remain poorly understood. Here we demonstrate the molecular basis for specific recognition of H3(1–15)K4me3 (residues 1–15 of histone H3 trimethylated at K4) by a plant homeodomain (PHD) finger of human BPTF (bromodomain and PHD domain transcription factor), the largest subunit of the ATP-dependent chromatin-remodelling complex, NURF (nucleosome remodelling factor). We report on crystallographic and NMR structures of the bromodomain-proximal PHD finger of BPTF in free and H3(1–15)K4me3-bound states. H3(1–15)K4me3 interacts through anti-parallel  $\beta$ -sheet formation on the surface of the PHD finger, with the long side chains of arginine 2 (R2) and K4me3 fitting snugly in adjacent pre-formed surface pockets, and bracketing an invariant tryptophan. The observed stapling role by non-adjacent R2 and K4me3 provides a molecular explanation for H3K4me3 site specificity. Binding studies establish that the BPTF PHD finger exhibits a modest preference for K4me3- over K4me2-containing H3 peptides, and discriminates against monomethylated and unmodified counterparts. Furthermore, we identified key specificity-determining residues from binding studies of H3(1–15)K4me3 with PHD finger point mutants. Our findings call attention to the PHD finger as a previously uncharacterized chromatin-binding module found in a large number of chromatin-associated proteins.

The packaging of DNA within chromosomes, the orderly replication and distribution of chromosomes, the maintenance of genomic integrity, and the regulated expression of genes depend upon nucleosomal histone proteins<sup>8–11</sup>. Introduction of covalent histone modifications, ATP-dependent chromatin remodelling, and utilization of histone variants are three major mechanisms by which variation can be introduced into the chromatin fibre. Histones methylated at lysine residues constitute a key epigenetic indexing system, demarcating transcriptionally active from inactive chromatin domains in eukaryotic genomes<sup>1–3</sup>. In particular, H3K4me3 has been linked to transcriptional activation in a range of eukaryotic species<sup>5–7</sup>. Recently, the chromodomain protein CHD1 (refs 12–14) and WDR5, a WD40 repeat protein<sup>15</sup>, have been shown to target H3 methylated at K4.

Because the trimethylated, rather than the dimethylated, form of K4 in H3 marks the transcription start site of eukaryotic genes, efforts

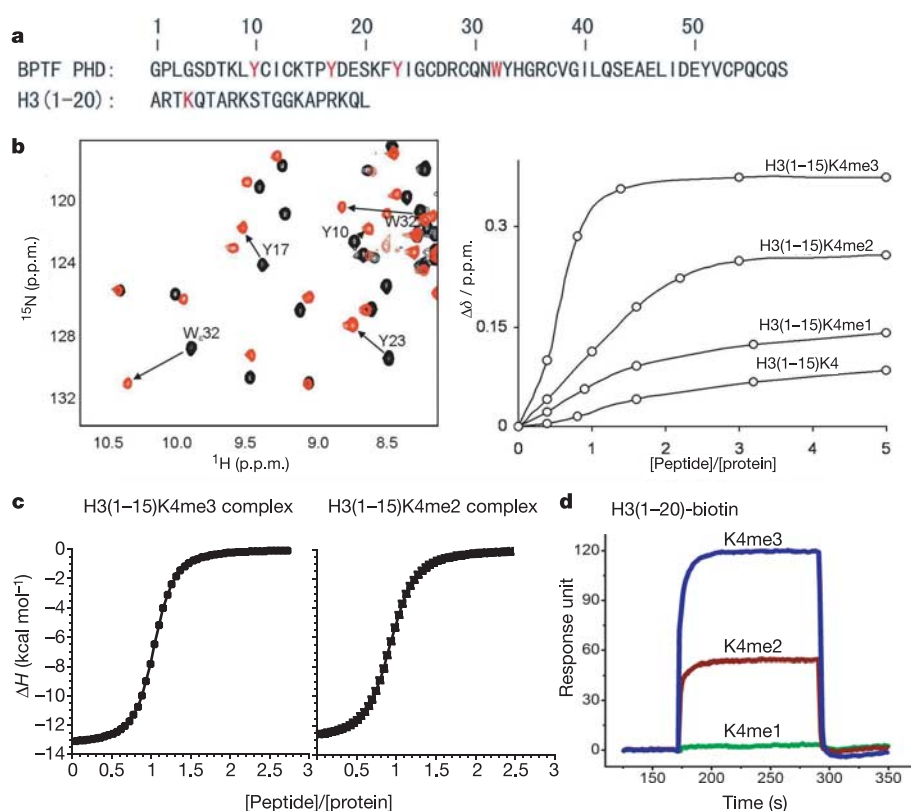
have been made to identify and characterize a H3K4me3 reader. An accompanying paper<sup>16</sup> reports on the discovery of trimethyl-specific H3K4me3 effectors: human BPTF and its orthologue, *Drosophila* NURF301, the largest subunit of NURF, the chromatin-remodelling complex that has been shown to facilitate transcriptional activation<sup>17–19</sup>.

Human BPTF and *Drosophila* NURF301 contain a bromodomain-proximal PHD finger (Fig. 1a and sequence alignment in Supplementary Fig. S1) involved in complex formation with H3 peptides trimethylated at K4 (ref. 16). NMR-based structures<sup>20–23</sup> of PHD domains have previously defined an interleaved, compact zinc finger scaffold containing a conserved Cys<sub>4</sub>HisCys<sub>3</sub> segment coordinated to two Zn<sup>2+</sup> ions. We observe well-resolved <sup>1</sup>H,<sup>15</sup>N-heteronuclear single quantum coherence (HSQC) spectra for the backbone amide region of <sup>15</sup>N-labelled human BPTF PHD finger in both the free state (Supplementary Fig. S2a) and in complex with H3(1–15)K4me3 peptide (Supplementary Fig. S2b). Several amide protons undergo shifts (up to 0.5 parts per million, p.p.m.) on complex formation (Supplementary Figs S3a and S4), including three tyrosines (Y10, Y17 and Y23) and one tryptophan (W32) (Fig. 1b, left panel). The chemical shift dependence of the amide resonance of Y17 as a function of added H3(1–15) peptides containing different K4 methylation states is plotted in Fig. 1b, right panel. These data establish that the BPTF PHD finger binds most tightly to H3(1–15)K4me3, somewhat weaker to H3(1–15)K4me2, and with the binding affinity decreasing further for the monomethylated and unmethylated states. Dissociation constants ( $K_d$ ) and binding enthalpies ( $\Delta H$ ) have been estimated from calorimetric measurements with  $K_d \approx 2.7 \mu\text{M}$  and  $\Delta H \approx -13.1 \text{ kcal mol}^{-1}$  for the H3K4me3 complex (Fig. 1c, left panel and Supplementary Fig. S5) and  $K_d \approx 5.0 \mu\text{M}$  and  $\Delta H \approx -12.1 \text{ kcal mol}^{-1}$  for the H3K4me2 complex (Fig. 1c, right panel and Supplementary Fig. S5). Binding has also been monitored using surface plasmon resonance with binding decreasing in the order K4me3 > K4me2 > K4me1 (Fig. 1d).

We have solved the 2.0 Å crystal structures (Supplementary Table S1) of a PHD finger-linker-bromodomain fragment of BPTF (sequence in Fig. 2a) in the free state (Fig. 2b) and in complex with H3(1–15)K4me3 peptide (Fig. 2c). The free (Fig. 2b) and bound (Fig. 2c) PHD finger-linker-bromodomain structures exhibit similar overall scaffolds (root mean square deviation (r.m.s.d.) of 0.76 Å for 165 C $\alpha$  atoms), with an  $\alpha$ -helical linker connecting the PHD finger and bromodomains. We have computed 1,030 Å<sup>2</sup> of buried surface area on complex formation.

We can readily and accurately trace the first six residues, A1-R2-T3-K4me3-Q5-T6, of the H3(1–15)K4me3 peptide in the complex (Fig. 2d and Supplementary Fig. S7a). This peptide segment is positioned on the surface of the BPTF finger (Figs 2e and 2f) through formation of an anti-parallel  $\beta$ -sheet with the K21 to D27 segment

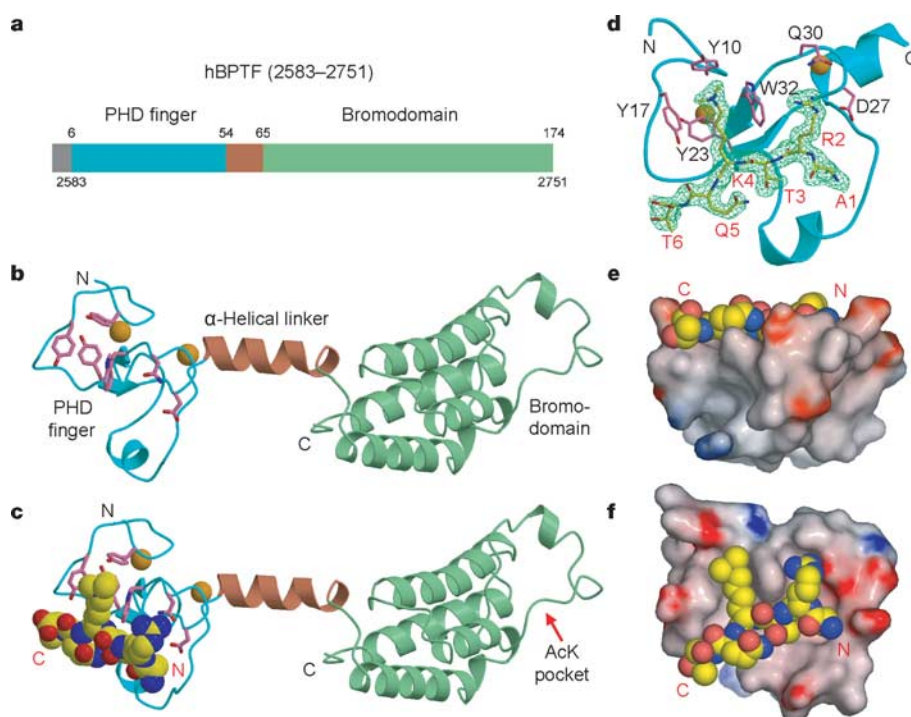
<sup>1</sup>Structural Biology Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA. <sup>2</sup>Laboratory of Chromatin Biology, Rockefeller University, New York, New York 10021, USA.



**Figure 1 | NMR, calorimetry and surface plasmon resonance-based studies of the binding of the BPTF PHD finger by H3(1–15)K4 peptides as a function of methylation state.** **a**, Sequences of the PHD finger of human BPTF and histone H3(1–20). **b**, Left panel: superposition of  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC NMR spectra showing BPTF PHD finger amide resonances in the absence (black) and presence (red) of 5 equivalents of H3(1–15)K4me3 peptide. Right panel: plots of chemical shift changes of the Y17 amide resonance of the BPTF PHD domain as a function of peptide concentration for different methylation states of H3K4. **c**, Isothermal titration calorimetry enthalpy plots for the binding of the BPTF PHD finger by H3(1–15)K4me3 (left panel) and H3(1–15)K4me2 (right panel) peptides. **d**, Surface plasmon resonance-based binding curves for the BPTF PHD finger with biotin-labelled H3(1–20) peptides as a function of K4 methylation state.

(Fig. 3a), stabilized by three direct hydrogen bonds between backbone amide and carbonyl groups involving the R2-T3-K4me3 segment. Notably, R2 and K4me3 are positioned in adjacent surface channels (Fig. 2f) separated by W32 (Fig. 2d, f), with this alignment facilitated by a hydrogen bond between the side chains of T3 and Q5 on the opposite side of the peptide backbone. The NMR complexation shifts are localized to the H3(1–15)K4me3-peptide-binding face of the PHD domain (Supplementary Fig. S3b).

The guanidinium group of R2 is directed towards G25 and is anchored in place through a number of direct and water-mediated hydrogen bonds involving its  $\text{N}_\epsilon$  and  $\text{N}_\eta$  protons and electrostatic interactions with D27 in the complex (Fig. 3b). The trimethyl group of K4me3 is positioned within a cage of four aromatic amino acids (Fig. 3c), of which Y23 forms the base and Y10, Y17 and W32 form three walls. The trimethyl ammonium group is stabilized by van der Waals and cation– $\pi$  interactions, similar to earlier results in the



**Figure 2 | Crystal structures of human BPTF PHD finger-linker-bromodomain in the free state and bound to H3(1–15)K4me3.** **a**, Domain architecture of BPTF bromodomain and proximal PHD finger. **b**, Ribbon representation of the crystal structure of the BPTF PHD finger-linker-bromodomain in the free state. Two bound Zn ions within the PHD fold are shown as balls. **c**, Crystal structure of the H3(1–15)K4me3-bound complex, with bound peptide shown in a space-filling representation. **d**, Structure of the PHD finger complex with the  $2F_o - F_c$  omit electron density ( $0.8\sigma$  level) highlighted for the bound H3(1–15)K4me3 peptide. **e**, Positioning of the H3(1–15)K4me3 peptide (space-filling representation) on the surface of the PHD finger portion (electrostatic surface representation with red as negatively charged and blue as positively charged surface) of the complex. **f**, Different view emphasizing the positioning of R2 and K4me3 in adjacent channels.



H3K9me3–HP1 chromodomain complex<sup>24–26</sup> (Supplementary Fig. S8a, b). There seem to be small conformational changes in the BPTF PHD finger on complex formation with H3(1–15)K4me3 peptide (stereo view in Fig. 3c).

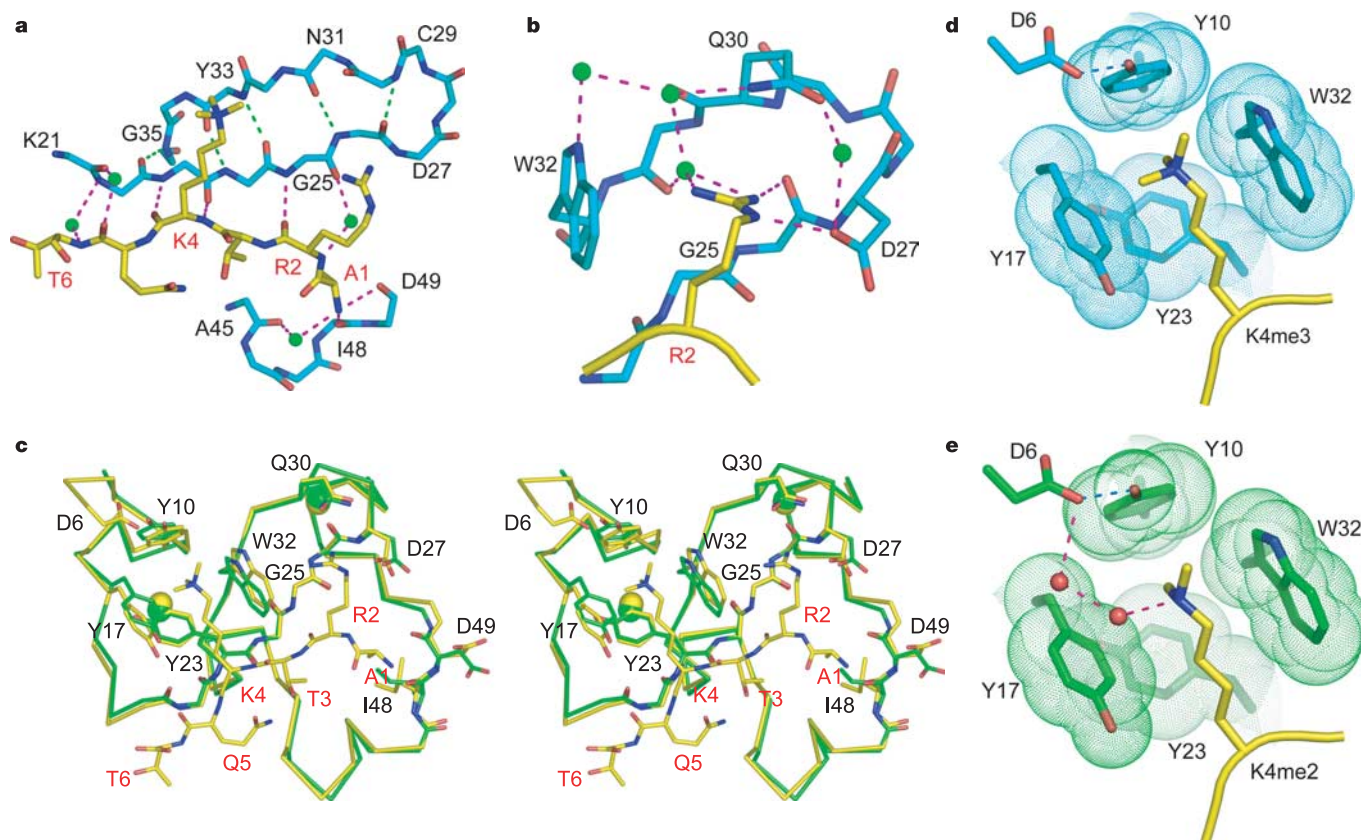
The structure of the complex also provides a likely explanation as to why the PHD finger of BPTF binds specifically to H3K4me3 but not to H3K9me3 or H3K27me3. The arrangement of W32-separated binding channels (Fig. 2d, f) requires that R2 and K4me3 be separated by one residue, as occurs for R2–T3–K4me3 but not for R8–K9me3 (Supplementary Fig. S8c, d) or R26–K27me3. We also note that the amino terminus of the peptide is anchored in position through a pair of hydrogen bonds from the NH<sub>2</sub> backbone to adjacent backbone carbonyls at the I48–D49 step (Fig. 3a).

We have also solved the crystal structure of the PHD finger–linker–bromodomain fragment of BPTF in complex with H3(1–15)K4me2 peptide at 1.9 Å resolution (Supplementary Table S1). There is very good superposition (r.m.s.d. = 0.12 Å for 176 C $\alpha$  atoms) of the crystal structures of the H3(1–15)K4me3-bound and H3(1–15)K4me2-bound complexes (stereo view in Supplementary Fig. S9). The dimethylammonium group of K4me2 of this complex (Fig. 3e) is positioned within the aromatic-amino-acid-lined cage of the PHD finger in an orientation similar to that observed for the trimethylammonium group of K4me3 in its complex (Fig. 3d). Importantly, the proton on the dimethylammonium group of K4me2 is linked through two bridging water molecules to the carboxylate of D6 (Fig. 3e; electron densities of bridging waters are shown in Supplementary Fig. S7b). The modest reduction in calorimetrically measured binding affinity and enthalpy from the K4me3 complex

to the K4me2 complex (Fig. 1c) might reflect the loss of  $\pi$ -stacking following replacement of a methyl group by a proton.

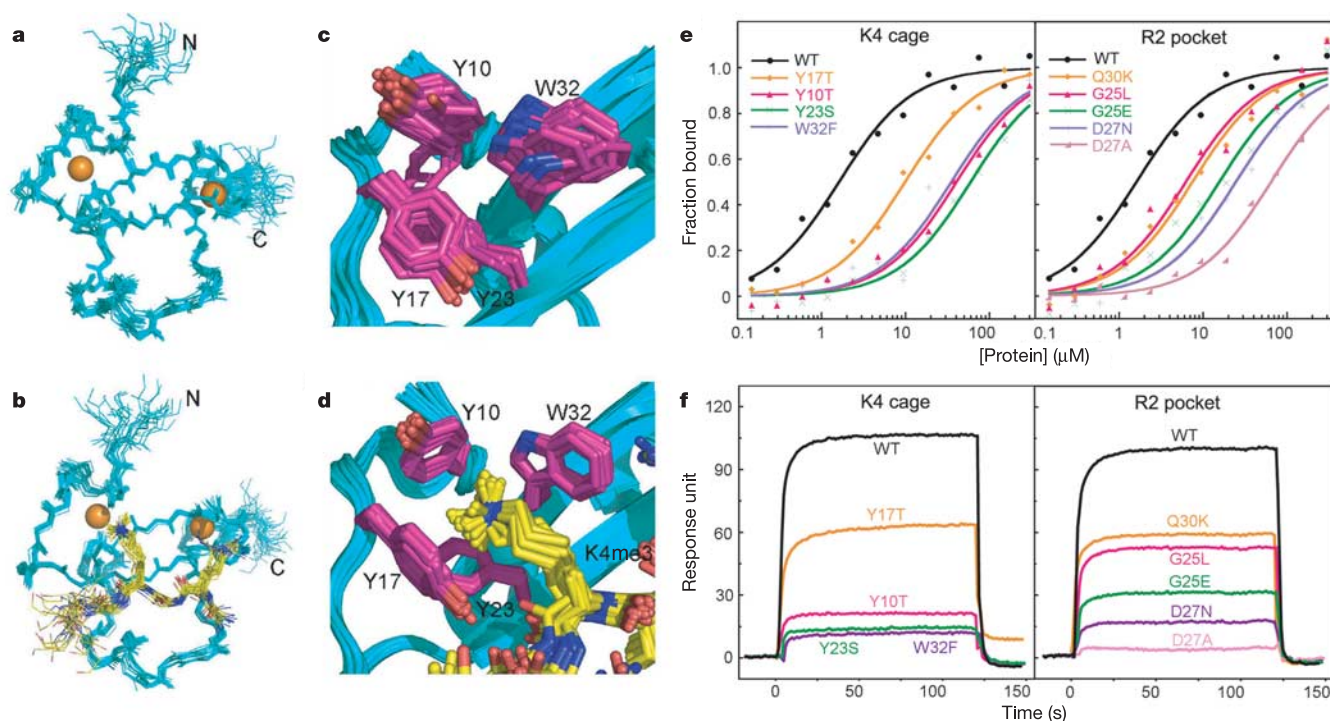
We have also solved the NMR structure of the BPTF PHD finger in the free state (superpositioned structures in Fig. 4a; statistics in Supplementary Table S2) and when bound to H3(1–15)K4me3 peptide (superpositioned structures in Fig. 4b and Supplementary Fig. S10; statistics are in Supplementary Table S2). There is excellent correspondence between the recognition features observed in the crystalline and solution states with r.m.s.d. values (51 C $\alpha$  backbone atoms) of 1.35 Å for the free BPTF PHD finger and 1.42 Å for the complex with bound H3(1–15)K4me3.

We have generated and purified point mutations of amino acids lining the K4me3-binding aromatic cage (Y10, Y17, Y23 and W32) and adjacent R2-binding channel (G25, D27 and Q30) of the BPTF PHD finger and estimated their impact on H3(1–15)K4me3 binding affinity by fluorescence polarization (Fig. 4e) and surface plasmon resonance (Fig. 4f) measurements. Fluorescence polarization measurements establish that aromatic cage mutants have an impact on binding affinity (Fig. 4e), with values dropping from  $K_d = 1.6 \pm 0.1 \mu\text{M}$  for wild-type to  $10.0 \pm 0.8 \mu\text{M}$  for Y17T,  $34.4 \pm 1.1 \mu\text{M}$  for W32F,  $40.3 \pm 4.5 \mu\text{M}$  for Y10T and  $59.7 \pm 8.3 \mu\text{M}$  for Y23S mutants. Thus, it seems that the aromatic cage residues, especially tryptophan at position 32, are critical for recognition. Y17, which is the least affected, varies the most among PHD finger homologues (Supplementary Fig. S11a). Corresponding reductions in binding affinity are also observed for R2-binding channel mutants, with values of  $7.5 \pm 0.4 \mu\text{M}$  for Q30K,  $6.2 \pm 0.3 \mu\text{M}$  for G25L,  $15.8 \pm 1.1 \mu\text{M}$  for G25E,  $24.6 \pm 0.6 \mu\text{M}$



**Figure 3** | Details of the intermolecular contacts in the H3(1–15)K4me3 peptide–BPTF PHD finger complex and comparison with its H3(1–15)K4me2-bound counterpart. **a**, Intermolecular backbone interactions between the A1–T6 segment of bound H3(1–15)K4me3 peptide and the PHD finger in the complex. **b**, Intermolecular hydrogen-bonding interactions involving the guanidinium group of R2 in the complex. **c**, Superposition of free (coloured green) and H3(1–15)K4me3-bound

complex (coloured yellow) of the BPTF PHD finger. **d**, **e**, Positioning of the trimethylated lysine of the H3(1–15)K4me3 peptide (**d**) and the dimethylated lysine of the H3(1–15)K4me2 peptide (**e**) within a four-aromatic-amino-acid cage of the BPTF PHD finger. Two bridging water molecules link the NH of K4me2 to the carboxylate of D6, as indicated by dashed lines.



**Figure 4 | NMR-based structures of free and H3(1–15)K4me3-bound BPTF PHD finger, together with binding studies of PHD finger mutants with H3(1–15)K4me3 peptide.** **a, b**, Backbone superposition of 20 NMR-based structures of the free BPTF PHD finger (residues 8–58) (**a**) and the H3(1–15)K4me3-bound BPTF PHD finger in solution (**b**). **c, d**, Superposition of structures centred about a cluster of aromatic residues (Y10, Y17, Y23 and W32) in the free BPTF PHD finger (**c**) and when bound

to H3(1–15)K4me3 (**d**). **e, f**, Fluorescence polarization-based binding curves for fluorescein-labelled H3(1–15)K4me3 peptide (**e**) and surface plasmon resonance-based binding curves for biotin-labelled H3(1–20)K4me3 peptide (**f**) with BPTF PHD finger mutants that cluster about either the K4-binding aromatic-amino-acid-lined cage (left panels) or the R2-binding pocket (right panels).

for D27N and  $59.2 \pm 3.4 \mu\text{M}$  for D27A mutants. Positions G25 and Q30 are also known to vary among PHD finger homologues (Supplementary Fig. S11a), whereas the more pronounced effect of D27 mutants suggests that this residue contributes critical electrostatic and hydrogen-bonding contacts within the R2 pocket. Surface plasmon resonance measurements yield a similar overall trend in reduced binding affinity (Fig. 4f).

The linker connecting the PHD and bromodomains of BPTF (Supplementary Fig. S12a) adopts an  $\alpha$ -helical conformation between residues 54–65 in both the free (Fig. 2b) and H3(1–15)K4me3 peptide-bound (Fig. 2c) states. This solvent-exposed  $\alpha$ -helix projects a hydrophilic surface with its N terminus fixed due to coordination involving the second zinc (stereo view in Supplementary Fig. S12b). Residues 66–71 adopt an extended conformation that forms extensive hydrogen bonds with elements of the bromodomain, with further stabilization associated with burial of L69, a conserved hydrophobic residue (Supplementary Fig. S12b).

Acetylated lysines in histone tails form complexes with bromodomains, an extensive family of protein modules (Supplementary Fig. S13a) found in histone acetyltransferases<sup>27,28</sup>. Structural studies of acetylated lysine-containing histone peptides in complex with bromodomains have enriched our understanding of the molecular basis of ligand selectivity<sup>29,30</sup>. The BPTF bromodomain superimposes well with the reported structure of yeast GCN5 bromodomain bound to H4K16ac peptide<sup>30</sup> (r.m.s.d. =  $0.88 \text{ \AA}$  for 100 C $\alpha$  atoms) (Supplementary Fig. S13b). The extent to which paired PHD finger-bromodomain in BPTF recognizes combinatorially methyl and acetyl marks in one or multiple tails, in keeping with the histone code hypothesis, remains to be determined.

Our work has provided the structural basis for substrate recognition by a tri- and dimethylated lysine-specific binding module: the PHD finger of BPTF/NURF301. The current working model,

supported by the functional data in an accompanying paper<sup>16</sup>, is that PHD finger-mediated binding of BPTF/NURF301 to H3K4me3 tails stabilizes the NURF complex on chromatin, resulting in NURF-catalysed nucleosome sliding and activation of *HOX* genes. As PHD fingers appear in a wide range of chromatin-associated proteins<sup>31</sup>, we anticipate that our complementary structural and functional studies will pave the way for new initiatives on these modification-binding modules and shed light on their role in a wide variety of proteins associated with chromatin, its remodelling and transcriptional regulation.

## METHODS

**Protein and peptide preparation.** The preparation and purification of the BPTF PHD finger and dual PHD finger-bromodomain, as well as isotope labelling with  $^{15}\text{N}$  and  $^{13}\text{C}$ ,  $^{15}\text{N}$  of the PHD finger for NMR studies and selenomethionine labelling of the dual PHD finger-bromodomain for X-ray studies, are described in detail in Supplementary Methods. Unmodified and methylated K4-modified H3 peptides and their carboxy-terminal fluorescein- and biotin-labelled counterparts were prepared and purified at our Institutional Core Facilities and their methylation states confirmed by mass spectroscopy.

**NMR assignments, complex formation and structure determination.** NMR experiments and data processing that result in backbone and side-chain resonance assignments of  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labelled BPTF PHD finger in the absence and presence of H3(1–15)K4me3 peptide, and structure determination computations for the PHD finger in free and H3(1–15)K4me3 peptide-bound states (statistics provided in Supplementary Table S2) are described in detail in Supplementary Information.

The NMR titration protocols for monitoring complex formation between the BPTF PHD finger and H3 K4 peptides as a function of methylation state are described in detail in Supplementary Information.

**Crystallization, data collection and structure determination.** Crystallization conditions, MAD data set collection of the free Se-Met-labelled BPTF PHD finger-chromodomain, as well as its complex with H3(1–15)K4me3 and H3(1–15)K4me2



peptides, and processing of data sets are described in detail in Supplementary Information.

The free form crystal belongs to space group  $P2_12_12_1$ . MAD phasing and initial automatic model building were performed with the program Solve/Resolve, followed by structure refinement and model rebuilding using the programs CNS and O, respectively. The model of free protein was determined at 2.0 Å against the Se-MAD peak data set with final  $R_{\text{work}}/R_{\text{free}}$  of 21.7/23.4. The H3(1–15)K4me3 and H3(1–15)K4me2 complex crystals belong to space group  $P2_1$  with three molecules in one asymmetric unit. The structure of the H3(1–15)K4me3 complex was solved by molecular replacement using Molrep in the CCP4 software package with the free protein as a search model. The model was refined to  $R_{\text{work}}/R_{\text{free}}$  19.8/22.7 at 2.0 Å by CNS. The structure of the H3(1–15)K4me2 complex was solved by difference Fourier synthesis using the CNS program with the H3(1–15)K4me3 complex structure as the starting model. The model was refined to  $R_{\text{work}}/R_{\text{free}}$  of 20.0/22.0 at 1.9 Å by CNS. Detailed refinement statistics are summarized in Supplementary Table S1.

**Additional protocols.** The protocols for fluorescence polarization, surface plasmon resonance and isothermal titration calorimetry measurements of the energetics of complex formation between H3K4me3 peptide as a function of methylation state and PHD mutants are described in detail in Supplementary Information.

Received 24 February; accepted 11 April 2006.

Published online 21 May 2006.

- Kouzarides, T. Histone methylation in transcriptional control. *Curr. Opin. Genet. Dev.* **12**, 198–209 (2002).
- Sims, R. J. III, Nishioka, K. & Reinberg, D. Histone lysine methylation: a signature for chromatin function. *Trends Genet.* **19**, 629–639 (2003).
- Martin, C. & Zhang, Y. The diverse functions of histone lysine methylation. *Nature Rev. Mol. Cell Biol.* **6**, 838–849 (2005).
- Fischle, W., Wang, Y. & Allis, D. Binary switches and modification cassettes in histone biology and beyond. *Nature* **425**, 475–479 (2003).
- Santos-Rosa, H. *et al.* Active genes are tri-methylated at K4 of histone H3. *Nature* **419**, 407–411 (2002).
- Santos-Rosa, H. *et al.* Methylation of histone H3 K4 mediates association of the Isw1 ATPase with chromatin. *Mol. Cell* **12**, 1325–1332 (2003).
- Schneider, R. *et al.* Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. *Nature Cell Biol.* **6**, 73–77 (2004).
- Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41–45 (2000).
- Wu, J. & Grunstein, M. 25 years after the nucleosome model: chromatin modifications. *Trends Biochem. Sci.* **25**, 619–623 (2000).
- Jenuwein, T. & Allis, C. D. Translating the histone code. *Science* **293**, 1074–1080 (2001).
- Felsenfeld, G. & Groudine, M. Controlling the double helix. *Nature* **421**, 448–453 (2003).
- Pray-Grant, M. G. *et al.* Chd1 chromodomain links histone H3 methylation with SAGA-SLIK-dependent acetylation. *Nature* **433**, 434–438 (2005).
- Sims, R. J. III *et al.* Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains. *J. Biol. Chem.* **280**, 41789–41792 (2005).
- Flanagan, J. F. *et al.* Double chromodomains cooperate to recognize the methylated histone H3 tail. *Nature* **438**, 1181–1185 (2005).
- Wysocka, J. *et al.* WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell* **121**, 859–872 (2005).
- Wysocka, J. *et al.* A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature advance online publication*, doi:10.1038/nature04815 (21 May 2006).
- Tsukiyama, T. & Wu, C. Purification and properties of an ATP-dependent nucleosome remodeling factor. *Cell* **83**, 1011–1020 (1995).
- Xiao, H. *et al.* Dual functions of largest NURF subunit NURF301 in nucleosome sliding and transcription factor interactions. *Mol. Cell* **8**, 531–543 (2001).
- Narlikar, G. J., Fan, H. Y. & Kingston, R. E. Cooperation between complexes that regulate chromatin structure and transcription. *Cell* **108**, 475–487 (2002).
- Pascual, J., Martinez-Yamout, M., Dyson, H. J. & Wright, P. E. Structure of the PHD zinc finger from human Williams-Beuren syndrome transcription factor. *J. Mol. Biol.* **304**, 723–729 (2000).
- Capili, A. D., Schultz, D. C., Rauscher, F. J. III & Borden, K. L. Solution structure of the PHD domain from the KAP-1 corepressor: structural determinants for PHD, RING and LIM zinc-binding domains. *EMBO J.* **20**, 165–177 (2001).
- Bottomley, M. J. *et al.* NMR structure of the first PHD finger of autoimmune regulator protein (AIRE1). *J. Biol. Chem.* **280**, 11505–11512 (2005).
- Kwan, A. H. *et al.* Engineering a protein scaffold from a PHD finger. *Structure* **11**, 803–813 (2003).
- Nielsen, P. R. *et al.* Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. *Nature* **416**, 103–107 (2002).
- Jacobs, S. A. & Khorasanizadeh, S. Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. *Science* **295**, 2080–2083 (2002).
- Min, J., Zhang, Y. & Xu, R.-M. Structural basis for specific binding of polycomb chromodomain to histone H3 methylated at Lys 27. *Genes Dev.* **17**, 1823–1828 (2003).
- Zeng, L. & Zhou, M. M. Bromodomain: an acetyl-lysine binding domain. *FEBS Lett.* **513**, 124–128 (2002).
- Bottomley, M. J. Structures of protein domains that create or recognize histone modifications. *EMBO Rep.* **5**, 464–469 (2004).
- Dhalluin, C. *et al.* Structure and ligand of a histone acetyltransferase bromodomain. *Nature* **399**, 491–496 (1999).
- Owen, D. J. *et al.* The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase Gcn5p. *EMBO J.* **19**, 6141–6149 (2000).
- Bienz, M. The PHD finger, a nuclear protein-interaction domain. *Trends Biochem. Sci.* **31**, 35–40 (2006).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** D.J.P. is supported by funds from the Abby Rockefeller Mauze Trust, and the Dewitt Wallace and Maloris Foundations. C.D.A. is supported by an NIH MERIT award and funds from Rockefeller University. S.I. holds a Ruth Kirschstein NIH postdoctoral fellowship and J.W. holds a Damon-Runyon CRF Fellowship. We thank the Peptide Core Facilities at Sloan-Kettering (S. S. Yi at Microchemistry and Proteomics) and Rockefeller University for the synthesis and purification of K4-methylated H3 peptides. We would like to thank the staff at beam line X25 at the Brookhaven National Laboratory and beam lines 23ID-D and 24ID-C of the Advanced Photon Source at the Argonne National Laboratory, supported by the US Department of Energy, for assistance with data collection. D.J.P. is a member of the New York Structural Biology Center, supported in part by funds from the NIH.

**Author Contributions** H.L. is responsible for the X-ray and surface plasmon resonance studies, S.I. for the NMR studies, W.W. for the calorimetric studies, B.D. for fluorescence polarization studies and J.W. for identification and functional characterization of the BPTF PHD finger as a H3K4me3 reader. D.J.P. and C.D.A. supervised the structural and functional (see companion paper) aspects of the project, respectively, and take overall responsibility for their joint research. All authors discussed the results and commented on the manuscript.

**Author Information** Coordinates of the X-ray structures of the BPTF PHD finger-linker-bromodomain in the free state and when bound to H3(1–15)K4me3 and H3(1–15)K4me2 peptides have been deposited in the RCSB Protein Data Bank under accession codes 2F6N, 2F6J and 2FSA, respectively. Coordinates of the NMR structures of the BPTF PHD finger in the free state and H3(1–15)K4me3 peptide-bound state have been deposited in the RCSB Protein Data Bank under accession codes 2FUI and 2FUU, respectively. Reprints and permissions information is available at [npg.nature.com/reprintsandpermissions](http://npg.nature.com/reprintsandpermissions). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.J.P. (pateld@mskcc.org) and C.D.A. (alliscd@rockefeller.edu).



## LETTERS

# ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression

Xiaobing Shi<sup>1</sup>, Tao Hong<sup>2</sup>, Kay L. Walter<sup>1</sup>, Mark Ewalt<sup>1</sup>, Eriko Michishita<sup>2</sup>, Tiffany Hung<sup>1</sup>, Dylan Carney<sup>1</sup>, Pedro Peña<sup>3</sup>, Fei Lan<sup>4</sup>, Mohan R. Kaadige<sup>5</sup>, Nicolas Lacoste<sup>6</sup>, Christelle Cayrou<sup>6</sup>, Foteini Davrazou<sup>3</sup>, Anjanabha Saha<sup>5</sup>, Bradley R. Cairns<sup>5</sup>, Donald E. Ayer<sup>5</sup>, Tatiana G. Kutateladze<sup>3</sup>, Yang Shi<sup>4</sup>, Jacques Côté<sup>6</sup>, Katrin F. Chua<sup>2,7</sup> & Or Gozani<sup>1</sup>

Dynamic regulation of diverse nuclear processes is intimately linked to covalent modifications of chromatin<sup>1,2</sup>. Much attention has focused on methylation at lysine 4 of histone H3 (H3K4), owing to its association with euchromatic genomic regions<sup>3,4</sup>. H3K4 can be mono-, di- or tri-methylated. Trimethylated H3K4 (H3K4me3) is preferentially detected at active genes, and is proposed to promote gene expression through recognition by transcription-activating effector molecules<sup>5</sup>. Here we identify a novel class of methylated H3K4 effector domains—the PHD domains of the ING (for inhibitor of growth) family of tumour suppressor proteins. The ING PHD domains are specific and highly robust binding modules for H3K4me3 and H3K4me2. ING2, a native subunit of a repressive mSin3a–HDAC1 histone deacetylase complex<sup>6</sup>, binds with high affinity to the trimethylated species. In response to DNA damage, recognition of H3K4me3 by the ING2 PHD domain stabilizes the mSin3a–HDAC1 complex at the promoters of proliferation genes. This pathway constitutes a new mechanism by which H3K4me3 functions in active gene repression. Furthermore, ING2 modulates cellular responses to genotoxic insults, and these functions are critically dependent on ING2 interaction with H3K4me3. Together, our findings establish a pivotal role for trimethylation of H3K4 in gene repression and, potentially, tumour suppressor mechanisms.

Binding of the ING2 PHD (ING2<sub>PHD</sub>) domain to the lipid signalling molecule phosphatidylinositol-5-phosphate (PtdIns(5)P) is critical for ING2 recruitment to chromatin during DNA damage responses<sup>7</sup>. Because some phosphatidylinositol phosphate (PtdInsP)-binding domains function through dual recognition of lipids and peptides<sup>8,9</sup>, we asked whether the ING2 PHD domain has protein, in addition to PtdInsP, ligands. In an *in vitro* pull-down screen for ING2<sub>PHD</sub> protein binding partners, we identified histone H3 (Supplementary Fig. S1a). ING2<sub>PHD</sub>-binding to H3 was independent of PtdIns(5)P-binding, and was confirmed in direct pull-down assays from purified bulk histones (Fig. 1a, c; Supplementary Fig. S1a).

The ING2 PHD domain bound robustly to native mononucleosomes purified from mammalian cells (Fig. 1b), but, in contrast to the ACF (ATP-dependent chromatin-assembly factor) and p300 (E1A binding protein p300) PHD domains, did not bind mononucleosomes reconstituted from recombinant histones (Supplementary Fig. S1b)<sup>10,11</sup>. Therefore, we postulated that this domain might recognize a post-translational H3 modification. Western blot analysis of ING2<sub>PHD</sub>-bound histones revealed significant methylation of H3 on lysine 4, but not lysine 9 (Fig. 1c). In *in vitro* binding assays of

methylated histone peptides, the ING2 PHD domain bound most strongly to H3K4me3, with lower affinity to H3K4me2, and not at all to numerous other peptides (Fig. 1d; Supplementary Fig. S1c).

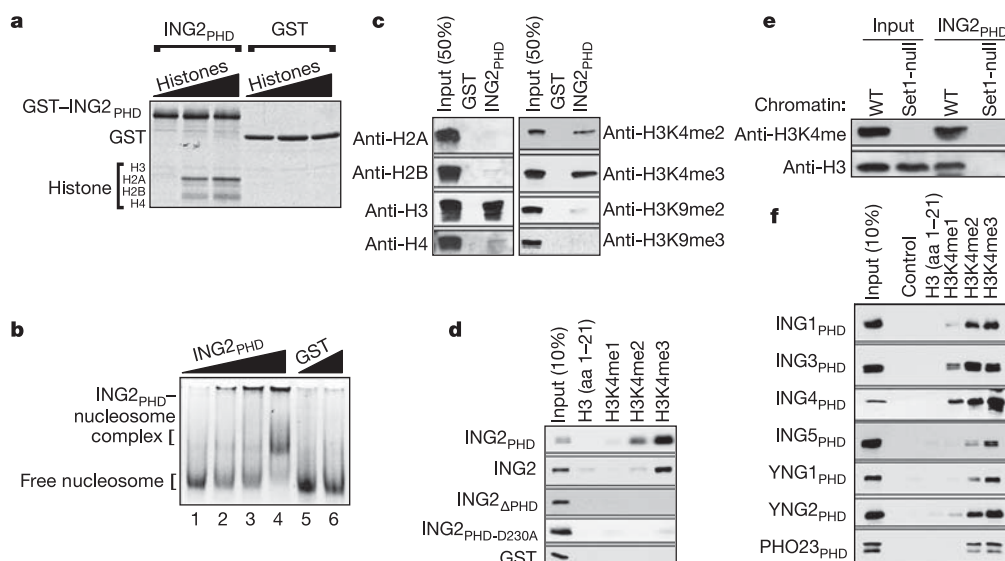
In the accompanying paper describing the crystal structure of the ING2<sub>PHD</sub>–H3K4me3 peptide complex, aspartate 230 (D230) of ING2 is implicated in H3K4me3 recognition<sup>12</sup>. Accordingly, substitution of D230 to alanine (ING2<sub>PHD-D230A</sub>) mostly abolished H3K4me3 peptide binding (Fig. 1d). The effect of this substitution is not due to unfolding of the PHD domain, as determined by NMR two-dimensional spectra<sup>12</sup>. Notably, this mutation has no effect on the lipid-binding activity of the ING2 PHD domain (Supplementary Fig. S2), and therefore allows selective abrogation of the ING2–H3K4me3 interaction for functional studies (see below). Finally, full-length ING2, but not a PHD domain deletion mutant (ING2<sub>ΔPHD</sub>), bound specifically to methylated H3K4 peptides (Fig. 1d). Thus, the ING2 PHD domain is necessary and sufficient for ING2 interaction with H3K4me3 *in vitro*.

The association of the ING2 PHD domain with H3 was increased by hypermethylation of bulk histones with the H3K4 methyltransferase SET7 (also known as SET9; ref. 13) and decreased by the H3K4 demethylase LSD1 (also known as AOF2; ref. 14) (Supplementary Fig. S3). Moreover, the ING2 PHD domain bound efficiently to H3 in chromatin purified from wild-type *Saccharomyces cerevisiae* strains, but not from strains lacking the H3K4 methyltransferase Set1 (Fig. 1e). Together, these data indicate that H3K4 methylation is critical for ING2<sub>PHD</sub>-binding to histone H3 at chromatin *in vitro*.

Like the ING2 PHD domain, the PHD domains of the *S. cerevisiae* and other human ING family members all bound preferentially to di- and tri-methylated H3K4, whereas that of the Mix2 autoantigen bound to trimethylated H3K36 (Fig. 1f; Supplementary Fig. S4a, b). Thus, recognition of methylated histone peptides is a general feature of at least a subset of PHD domains, with the ING PHD family constituting a novel class of binding modules for methylated H3K4.

The functional role of ING2 in the context of mSin3a–HDAC1 complexes is unclear<sup>6</sup>. We postulated that ING2 bridges the mSin3a–HDAC1 complex with methylated H3K4 through its PHD domain, and thereby promotes deacetylation of nearby acetylated histone residues. To test this idea, we purified ING2–mSin3a–HDAC1 complexes containing ING2, ING2<sub>ΔPHD</sub> or ING2<sub>D230A</sub>. Neither mutation altered the composition of the ING2 complexes (Fig. 2a; Supplementary Fig. S5). In peptide pull-down assays, the wild-type, but not mutant, ING2 complexes bound preferentially to H3K4me3 (Fig. 2b). Furthermore, hypermethylation of H3K4 by SET7 enhanced the histone deacetylase activity of the wild-type, but not mutant, ING2

<sup>1</sup>Department of Biological Sciences, Stanford University, and <sup>2</sup>Department of Medicine, Stanford University School of Medicine, Stanford, California 94305, USA. <sup>3</sup>Department of Pharmacology, University of Colorado Health Sciences Center, Aurora, Colorado 80045, USA. <sup>4</sup>Department of Pathology, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>5</sup>Huntsman Cancer Institute and Department of Oncological Sciences, University of Utah, Salt Lake City, Utah 84112, USA. <sup>6</sup>Laval University Cancer Research Center, Quebec City, Quebec G1R 2J6, Canada. <sup>7</sup>Geriatric Research, Education and Clinical Center, VA Palo Alto Health Care System, Palo Alto, California 94304, USA.



**Figure 1 | The ING2 PHD domain specifically binds to H3K4me3 *in vitro*.** **a**, Coomassie-blue stain of glutathione *S*-transferase (GST)–ING2<sub>PHD</sub> and GST control pull downs from calf thymus histones. **b**, Gel shift showing ING2<sub>PHD</sub>-binding to purified native nucleosomes. Ethidium-bromide stain of nucleosomal DNA on a non-denaturing gel. **c**, Western blot analysis of GST pull-down assays as in **a**. **d**, The ING2 PHD domain preferentially binds H3K4me3 peptides. Western analysis of histone peptide pull downs with the

indicated GST fusion proteins and biotinylated peptides. Peptide integrity was confirmed with known methyl-lysine-binding domains (Supplementary Fig. S1d, e). aa, amino acids. **e**, Methylation at H3K4 is required for ING2<sub>PHD</sub>-binding to chromatin *in vitro*. Western analysis of GST–ING2<sub>PHD</sub> pull-down assays of chromatin purified from wild-type or Set1-null *S. cerevisiae* strains. **f**, PHD domains of yeast and human ING family proteins preferentially bind to H3K4me2/3 in histone peptide binding assays.

complexes (Figs 2c, d). Therefore, recognition of methylated H3K4 by the ING2 PHD domain promotes the histone deacetylase activity of mSin3a–HDAC1 complexes *in vitro*.

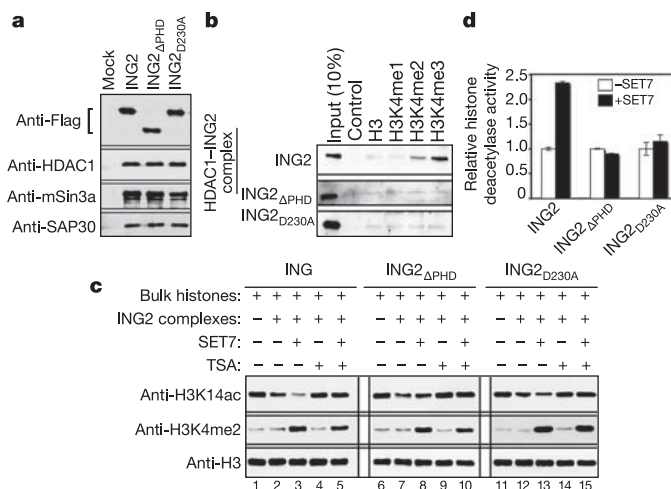
We next studied the recognition of methylated H3K4 by the ING2 PHD domain in the context of physiological chromatin in cells. In protein–protein chromatin immunoprecipitation (ChIP) assays<sup>15</sup>, wild-type ING2 associated strongly with H3K4me3, but only weakly with H3K4me2 and not at all with other modified histones, suggesting that H3K4me3 is the preferred *in vivo* target of ING2 (Fig. 3a; data not shown). Moreover, the H3K4me3 interaction was mostly abolished by the D230A mutation and several additional substitutions that disrupt H3K4me3-binding *in vitro*<sup>12</sup> (Fig. 3a, b; data not shown). We conclude that the PHD domain is required for the association of ING2 with H3K4me3 at chromatin *in vivo*.

To determine whether endogenous H3K4me3 is important for ING2 association with H3, we knocked down the expression of the methyltransferase component, the WD40-repeat protein WDR5, which leads to a specific reduction in H3K4me3 levels (Fig. 3c)<sup>16</sup>. This was accompanied by substantially reduced binding of both Flag-tagged and endogenous ING2 to H3 at chromatin (without affecting complex formation with HDAC1) (Fig. 3d, e). Thus, H3K4me3 is required for interaction of ING2 with H3, and endogenous ING2 specifically associates with trimethylated H3K4 at chromatin *in vivo*.

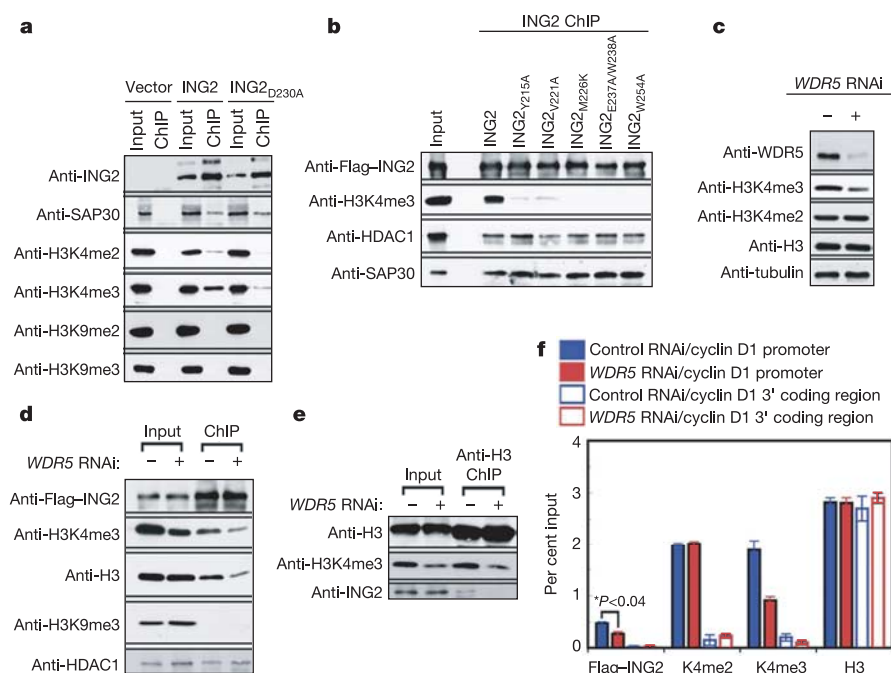
To investigate ING2 association with chromatin at a target gene, we focused on the cell-cycle regulator cyclin D1. This gene is transcriptionally inactivated by the mSin3a–HDAC1 complex<sup>17</sup>. In ChIP analyses, Flag–ING2 (and tri- and di-methylated H3K4, as expected<sup>3,4</sup>) are specifically detected at the cyclin D1 promoter but not at 3' coding regions (Fig. 3f; Supplementary Fig. S6a). Furthermore, WDR5 knockdown decreased ING2 occupancy at the cyclin D1 promoter (Fig. 3f). Thus, trimethylation of H3K4 is important for ING2 association at the cyclin D1 promoter.

To investigate the biological significance of the ING2–H3K4me3 interaction, we used an ING2 complementation system<sup>7</sup>, whereby RNA interference (RNAi)-resistant wild-type or mutant (D230A or ΔPHD) ING2 constructs were stably expressed in ING2 knockdown cells. ING2

expression was substantially reduced in the knockdown cells, and the different ING2 constructs reconstituted expression at comparable levels (Fig. 4a). Notably, the ING2 knockdown cells showed impaired cyclin D1 mRNA repression in response to doxorubicin compared



**Figure 2 | Methylated H3K4 recognition by the ING2 PHD domain enhances ING2-associated HDAC1 histone deacetylase activity *in vitro*.** **a**, Western analysis of affinity-purified wild-type and mutant Flag–ING2 complexes. Mock, empty vector control immunoprecipitation. **b**, Binding of ING2 complexes to methylated H3K4 peptides requires an intact PHD domain. Anti-ING2 western analysis of histone peptide pull downs from the indicated HDAC1–ING2 complexes. **c**, Histone deacetylation by HDAC1 in wild-type, but not mutant, ING2 complexes is increased by binding of the ING2 PHD domain to methylated H3K4. Western analysis of *in vitro* histone deacetylation reactions by ING2 complexes in the presence or absence of methylation by SET7. Histone deacetylase inhibition with trichostatin A (TSA) is shown as a control. **d**, Quantification of relative histone deacetylase activity of ING2 complexes as in **c** from three independent experiments. Error bars indicate the s.e.m.

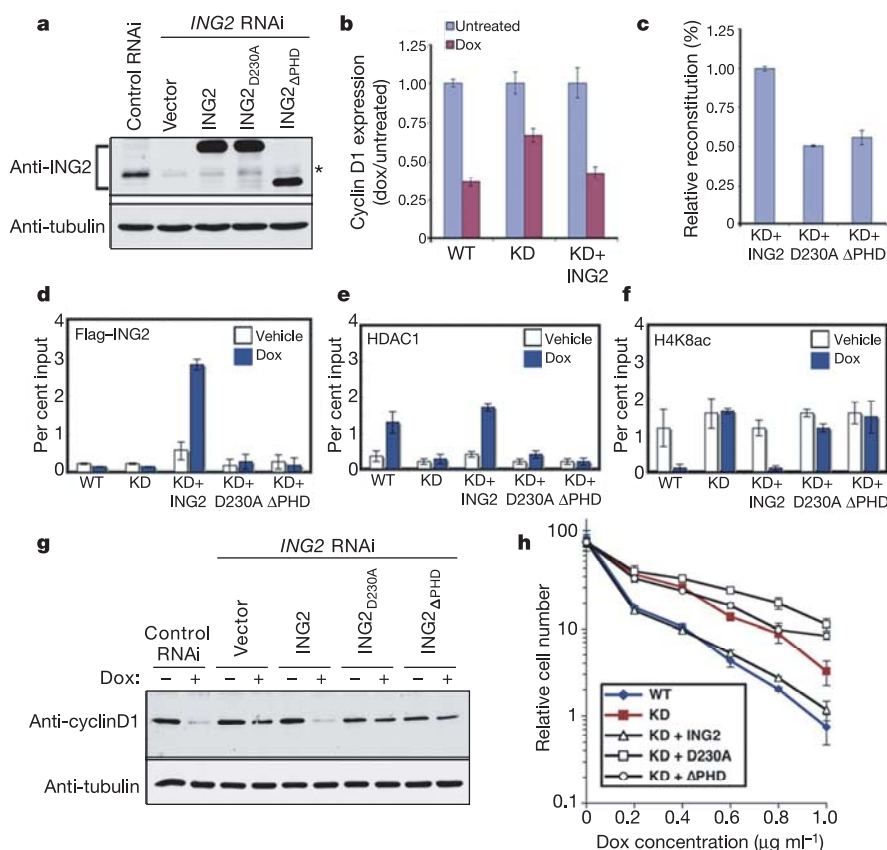


**Figure 3 | The ING2 interaction with trimethylated H3K4 occurs *in vivo* and requires an intact PHD domain.** **a, b,** *In vivo* association of wild-type, but not PHD domain mutant, ING2 proteins with H3K4me3 at chromatin. Western analysis of proteins crosslinked *in situ* to wild-type or mutant Flag-ING2 proteins in protein–protein ChIPs. Note equal binding of mutant and wild-type ING2 to SAP30 (**a**) and HDAC1 (**b**). Input loaded represents 5% of total. **c,** Decreased H3K4me3 levels following WDR5 knockdown. Western analysis of HEK293T cells transfected with WDR5 or control RNA interference (RNAi) vectors. **d,** Flag-ING2 association with endogenous H3 requires H3K4me3. Western analysis of protein–protein ChIPs as in **a** with or without WDR5 knockdown. **e,** Endogenous association of H3 and ING2 requires H3K4me3. Western analysis of H3-bound proteins by protein–protein ChIP, with or without WDR5 knockdown. **f,** WDR5 knockdown decreases ING2 occupancy at the cyclin D1 promoter. Levels (per cent input = ChIP/input  $\times$  100) of cyclin D1 promoter and 3' coding region sequences in the indicated ChIPs, with or without WDR5 knockdown. Error bars indicate the s.e.m. from three independent experiments.

with control (wild-type) cells, and complementation of knockdown cells with wild-type ING2 mostly restored efficient cyclin D1 repression (Fig. 4b). In contrast, the ING2<sub>D230A</sub> and ING2<sub>ΔPHD</sub> proteins showed significantly reduced reconstitution activities (Fig. 4c). Thus, PHD domain recognition of H3K4me3 is important for ING2-mediated repression of cyclin D1 mRNA expression in response to DNA damage.

Next, we examined promoter occupancy of the ING2–HDAC1

complexes in response to DNA damage. Doxorubicin treatment increased promoter occupancy of wild-type, but not mutant, ING2 at the cyclin D1 promoter (Fig. 4d), as well as a second HDAC1-regulated DNA-damage response gene, *c-Myc* (Supplementary Fig. S7)<sup>18</sup>. Furthermore, HDAC1 occupancy and histone deacetylation at the cyclin D1 promoter both increased on DNA damage, and these responses, which were largely abrogated in the knockdown cells, were restored by reconstitution with wild-type, but not mutant,



**Figure 4 | Recognition of H3K4me3 by ING2 PHD domain *in vivo* is critical for ING2 function.** **a,** Western analysis of ING2 expression in the indicated cell lines. \*, endogenous ING2. **b,** ING2 is required for repression of cyclin D1 expression in response to doxorubicin (dox). Quantitative PCR of cyclin D1 mRNA, with or without doxorubicin (0.2  $\mu\text{g ml}^{-1}$ , 24 h), in wild-type (“WT”), knockdown (“KD”) and knockdown reconstituted with ING2 (“KD+ING2”) cell lines. **c,** Impaired activity of mutant ING2 proteins in reconstitution of doxorubicin-dependent cyclin D1 repression, as in **b**, relative to wild-type ING2 (“KD+ING2”). In **b** and **c**, error bars indicate the s.e.m. of triplicate experiments; *P*-values < 0.01. **d–f,** DNA-damage-dependent increase of ING2–HDAC1 occupancy at the cyclin D1 promoter requires methylated H3K4-binding. ChIPs with antibodies to Flag (**d**), HDAC1 (**e**) and H4K8ac (**f**) at the cyclin D1 promoter in the indicated cell lines, with or without doxorubicin (2  $\mu\text{M}$ , 1 h). Error bars indicate the s.e.m. of at least three independent experiments. *P*-values < 0.05. **g,** Western analysis of cyclin D1 protein in the indicated cell lines with or without doxorubicin (0.2  $\mu\text{g ml}^{-1}$ , 24 h). **h,** Impaired doxorubicin sensitivity following ING2 knockdown or reconstitution with H3K4me3-binding-defective ING2. Relative cell viability is normalized to untreated cells. Error bars indicate the s.e.m. of 3–6 independent experiments.



ING2 (Fig. 4e, f). H3K4me2, H3K4me3 and H3 levels were uniform across the different cell lines and not altered by doxorubicin treatment (Supplementary Fig. S6b). On western analysis, cyclin D1 protein levels were repressed following DNA damage in the presence of wild-type ING2 (in both control cells and knockdown cells reconstituted with ING2), but only slightly reduced in the knockdown cells or knockdown cells reconstituted with mutant ING2 proteins (Fig. 4g). We conclude that recognition of H3K4me3 by the ING2 PHD domain is critical for proper occupancy of the ING2–HDAC1 complex at target promoters during DNA damage responses and active transcriptional repression of the promoter's cognate gene.

Finally, H3K4me3 recognition by the PHD domain is critical for ING2 regulation of cellular functions. Specifically, ING2 knockdown significantly impairs cellular sensitivity to doxorubicin (Fig. 4h), and reconstitution of knockdown cells with wild-type, but not mutant, ING2 reverses this effect (Fig. 4h). Together, these data argue that the H3K4me3-binding activity of ING2 is critical for ING2 function in the cellular response to DNA damage.

Our findings and previous work demonstrate that the PHD domain of ING2 is a dual-specificity module that binds both PtdIns(5)P and H3K4me3. These activities are separable by specific mutations within the domain, and both seem critical for ING2 cellular functions. It is possible that PtdIns(5)P-binding could mediate trafficking of ING2 to target promoters, where binding to H3K4me3 then contributes to retention of ING2 at these promoters (see Supplementary Fig. S8). In this regard, the multi-functionality of the ING2 PHD domain may contribute to integrating complex nuclear signalling events that link DNA damage to chromatin responses.

Methylation of H3K4 is proposed to have a pivotal role in gene activation by serving as a binding platform for different transcription-promoting factors<sup>16,19–21</sup>. Our findings identify the ING2 PHD domain as the first effector domain for methylated H3K4 that links this modification to transcriptional repression. By focusing HDAC1 repressor complexes on actively transcribed genes, recognition of H3K4me3 by ING2 may be important for the efficiency of acute gene repression. Such a mechanism may be particularly important in the context of cellular responses to acute stress, such as DNA damage, in which rapid shut-off of proliferation genes is critical to prevent propagation of cells harbouring damaged DNA. The PHD domains of the other ING proteins also bind methylated H3K4, and thus may link this modification to additional cellular functions. Moreover, the PHD domain of the nucleosome remodelling factor (NURF) complex component BPTF (bromodomain and PHD domain transcription factor) also recognizes methylated H3K4 (ref. 22), suggesting a more general role for PHD domains as methyl-lysine effector domains. Together, our findings highlight the notion that the recognition of chromatin modifications by effector proteins, rather than the specific modification *per se*, determines biological function, and, as such, greatly expands the diversity of signalling at chromatin.

## METHODS

**Histone peptide binding assays.** Biotinylated histone peptides were synthesized at Stanford Protein and Nucleic Acid facility or purchased from Upstate Biotechnology. Briefly, 0.5 µg of peptides were incubated with 1 µg of protein in binding buffer (50 mM Tris-HCl, pH 7.5, 300 mM NaCl, 0.1% (v/v) NP-40, 1 mM phenylmethylsulphonyl fluoride (PMSF)) overnight at 4 °C. After 1 h of incubation, streptavidin beads (Amersham) were washed three times and subjected to western analysis.

Protein–protein ChIP assays for detection of *in situ* ING2–histone interactions was performed as described<sup>15</sup>. Cell viability assays were performed as described<sup>23</sup>. Information about antibodies, recombinant vectors, RNAi

sequences and primers are available in Supplementary Information and on request. Detailed methods can be found in Supplementary Information.

Received 27 February; accepted 17 April 2006.

Published online 21 May 2006.

- Jenuwein, T. & Allis, C. D. Translating the histone code. *Science* **293**, 1074–1080 (2001).
- Kurdiani, S. K. & Grunstein, M. Histone acetylation and deacetylation in yeast. *Nature Rev. Mol. Cell Biol.* **4**, 276–284 (2003).
- Schneider, R. *et al.* Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. *Nature Cell Biol.* **6**, 73–77 (2004).
- Bernstein, B. E. *et al.* Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120**, 169–181 (2005).
- Bannister, A. J. & Kouzarides, T. Histone methylation: recognizing the methyl mark. *Methods Enzymol.* **376**, 269–288 (2004).
- Doyon, Y. *et al.* ING tumor suppressor proteins are critical regulators of chromatin acetylation required for genome expression and perpetuation. *Mol. Cell* **21**, 51–64 (2006).
- Gozani, O. *et al.* The PHD finger of the chromatin-associated protein ING2 functions as a nuclear phosphoinositide receptor. *Cell* **114**, 99–111 (2003).
- Zimmermann, P. *et al.* PIP<sub>2</sub>–PDZ domain binding controls the association of syntrophin with the plasma membrane. *Mol. Cell* **9**, 1215–1225 (2002).
- Stolt, P. C. *et al.* Origins of peptide selectivity and phosphoinositide binding revealed by structures of disabled-1 PTB domain complexes. *Structure (Camb.)* **11**, 569–579 (2003).
- Ragvin, A. *et al.* Nucleosome binding by the bromodomain and PHD finger of the transcriptional cofactor p300. *J. Mol. Biol.* **337**, 773–788 (2004).
- Eberharter, A., Vetter, I., Ferreira, R. & Becker, P. B. ACF1 improves the effectiveness of nucleosome mobilization by ISWI through PHD–histone contacts. *EMBO J.* **23**, 4029–4039 (2004).
- Peña, P. V. *et al.* Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. *Nature* advance online publication, doi:10.1038/nature04814 (21 May 2006).
- Wang, H. *et al.* Purification and functional characterization of a histone H3-lysine 4-specific methyltransferase. *Mol. Cell* **8**, 1207–1217 (2001).
- Shi, Y. *et al.* Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**, 941–953 (2004).
- Ricke, R. M. & Bielinsky, A. K. Easy detection of chromatin binding proteins by the Histone Association Assay. *Biol. Proced. Online* **7**, 60–69 (2005).
- Wysocka, J. *et al.* WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell* **121**, 859–872 (2005).
- Dannenberg, J. H. *et al.* mSin3A corepressor regulates diverse transcriptional networks governing normal and neoplastic growth and survival. *Genes Dev.* **19**, 1581–1595 (2005).
- Ho, J. S., Ma, W., Mao, D. Y. & Benchimol, S. p53-dependent transcriptional repression of c-myc is required for G<sub>1</sub> cell cycle arrest. *Mol. Cell Biol.* **25**, 7423–7431 (2005).
- Santos-Rosa, H. *et al.* Methylation of histone H3 K4 mediates association of the Isw1p ATPase with chromatin. *Mol. Cell* **12**, 1325–1332 (2003).
- Pray-Grant, M. G., Daniel, J. A., Schieltz, D., Yates, J. R. III & Grant, P. A. Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation. *Nature* **433**, 434–438 (2005).
- Sims, R. J. III *et al.* Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains. *J. Biol. Chem.* **280**, 41789–41792 (2005).
- Wysocka, J. *et al.* A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* advance online publication, doi:10.1038/nature04815 (21 May 2006).
- Mostoslavsky, R. *et al.* Genomic instability and aging-like phenotype in the absence of mammalian SIRT6. *Cell* **124**, 315–329 (2006).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank Y. Zhang for the SET7 expression vector, C. Harris for ING3, ING4 and ING5 complementary DNAs, J. Wysocka and W. Herr for a WDR5 antibody, J. Wysocka and C. D. Allis for communicating unpublished results, J. Yuan for ING2 antibodies and A. Sanchez for peptide synthesis. This work was supported by NIH grants to O.G., Y.S. and B.R.C. O.G. is a recipient of a Burroughs Wellcome Career Development Award in Biomedical Sciences.

**Author Information** Reprints and permissions information is available at [npg.nature.com/reprintsandpermissions](http://npg.nature.com/reprintsandpermissions). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to O.G. ([ogozani@stanford.edu](mailto:ogozani@stanford.edu)).

## LETTERS

# Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2

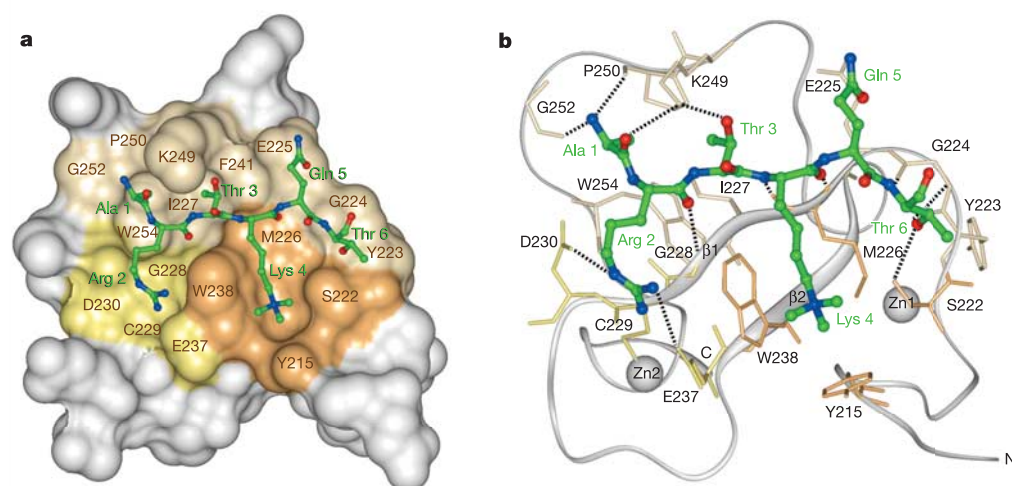
Pedro V. Peña<sup>1</sup>, Foteini Davrazou<sup>1</sup>, Xiaobing Shi<sup>3</sup>, Kay L. Walter<sup>3</sup>, Vladislav V. Verkhusha<sup>4</sup>, Or Gozani<sup>3</sup>, Rui Zhao<sup>2</sup> & Tatiana G. Kutateladze<sup>1</sup>

Covalent modifications of histone tails have a key role in regulating chromatin structure and controlling transcriptional activity. In eukaryotes, histone H3 trimethylated at lysine 4 (H3K4me3) is associated with active chromatin and gene expression<sup>1–4</sup>. We recently found that plant homeodomain (PHD) finger of tumour suppressor ING2 (inhibitor of growth 2) binds H3K4me3 and represents a new family of modules that target this epigenetic mark<sup>5</sup>. The molecular mechanism of H3K4me3 recognition, however, remains unknown. Here we report a 2.0 Å resolution structure of the mouse ING2 PHD finger in complex with a histone H3 peptide trimethylated at lysine 4. The H3K4me3 tail is bound in an extended conformation in a deep and extensive binding site consisting of elements that are conserved among the ING family of proteins. The trimethylammonium group of Lys 4 is recognized by the aromatic side chains of Y215 and W238 residues, whereas the intermolecular hydrogen-bonding and complementary surface interactions, involving Ala 1, Arg 2, Thr 3 and Thr 6 of the peptide, account for the PHD finger's high specificity and affinity. Substitution of the binding site residues disrupts H3K4me3 interaction *in vitro* and impairs the ability of ING2 to induce apoptosis *in vivo*. Strong binding of other ING and YNG PHD fingers suggests that the recognition of H3K4me3 histone code is a general feature of the ING/YNG proteins. Elucidation of the mechanisms underlying this novel function of PHD fingers provides a basis for deciphering the role of the ING family of tumour suppressors in chromatin regulation and signalling.

Specific post-translational modifications of histone tails facilitate

recruitment of distinct proteins to nucleosomes to either activate gene expression or condense the chromatin<sup>6</sup>. Despite a wide variety of known histone marks, only a few protein domains have been identified that recognize or 'read' the precise tail modifications. This includes bromodomain, which binds acetylated lysine residues of histone H4 (refs 7, 8), and chromodomain, which targets methylated lysine residues of histone H3 (refs 9–13). A number of recent reports suggest that PHD modules are involved in chromatin remodelling<sup>14,15</sup>. We found that the carboxy-terminal PHD finger of tumour suppressor ING2 directly associates with the histone H3 tail trimethylated at Lys 4. To understand the molecular mechanism of this histone code recognition, we determined a 2 Å resolution crystal structure of the mouse ING2 PHD finger bound to the H3K4me3 peptide and established the determinants of specificity and functional significance of this interaction.

The structure of the complex indicates a conserved mode of methyl-lysine recognition by PHD fingers of ING proteins and reveals key elements that define the binding specificity. The overall fold of the ING2 PHD finger in the complex is similar to the fold previously found in ligand-free PHD modules<sup>16</sup>. The structure consists of three loops stabilized by two zinc-binding clusters and a small double-stranded anti-parallel  $\beta$ -sheet (Fig. 1 a, b). The structure of the ING2 PHD finger in complex with H3K4me3 peptide superimposes well with the structure of the free PHD finger of identical ING1L protein (Protein Data Bank 1WES) with a root-mean-squared deviation of 0.8 Å, thus indicating that binding of the peptide does not induce significant conformational changes in the protein.



**Figure 1 | Structure of ING2 PHD finger in complex with a histone H3 peptide trimethylated at Lys 4.** **a**, The PHD finger is shown as a solid surface with the binding site residues coloured and labelled. The Lys 4 and Arg 2 binding grooves are in brown and yellow, respectively. The histone peptide is shown as a ball-and-stick model with C, O and N atoms coloured green, red and blue, respectively. **b**, Ribbon diagram of the structure. Dashed lines represent intermolecular hydrogen bonds.

<sup>1</sup>Department of Pharmacology, and <sup>2</sup>Department of Biochemistry and Molecular Genetics, University of Colorado Health Sciences Center, Aurora, Colorado 80045, USA.

<sup>3</sup>Department of Biological Sciences, Stanford University, Stanford, California 94305, USA. <sup>4</sup>Department of Anatomy and Structural Biology, Albert Einstein College of Medicine, Bronx, New York 10461, USA.

**Table 1 | Affinities of the wild-type and mutant PHD fingers for the histone tail peptides**

| PHD proteins           | Ligand        | $K_d$ ( $\mu$ M)         |
|------------------------|---------------|--------------------------|
| ING2 wild type         | H3K4me3       | $1.5 \pm 1^*$            |
| ING2 wild type         | H3K4me2       | $15 \pm 4^*$             |
| ING2 wild type         | H3K4me1       | $208 \pm 80^*$           |
| ING2 wild type         | H3K9me3       | $2,000 \pm 60^\dagger$   |
| ING2 wild type         | H3K9me2       | $2,320 \pm 300^*\dagger$ |
| ING2 wild type         | H3K9me1       | $2,380 \pm 800^\dagger$  |
| ING2 wild type         | H3 unmodified | $2,240 \pm 350^*\dagger$ |
| ING2 wild type         | H4K20me3      | $>10,000^\dagger$        |
| ING2 wild type         | H4K20me2      | $>10,000^*\dagger$       |
| ING2 wild type         | H4K20me1      | $>7,000^*\dagger$        |
| ING2 Y215A             | H3K4me3       | $>5,000^*$               |
| ING2 S222A             | H3K4me3       | $24 \pm 10^*$            |
| ING2 Y223A/E225A       | H3K4me3       | $326 \pm 169^*\dagger$   |
| ING2 D230A             | H3K4me3       | $51 \pm 15^*$            |
| ING2 E237A/W238A       | H3K4me3       | $726 \pm 50^*$           |
| ING2 K249A/K251A/K253A | H3K4me3       | $1.1 \pm 0.6^*$          |
| ING1                   | H3K4me3       | $3.3 \pm 1.6^*$          |
| ING3                   | H3K4me3       | $6.9 \pm 1.1^*$          |
| ING4                   | H3K4me3       | $7.9 \pm 2^*$            |
| ING5                   | H3K4me3       | $2.4 \pm 1^*$            |
| YNG1                   | H3K4me3       | $2.3 \pm 0.9^*$          |
| YNG2                   | H3K4me3       | $5.1 \pm 1.8^*$          |
| PHO23                  | H3K4me3       | $4.5 \pm 1.3^*$          |

 $K_d$ , dissociation constant.

\* Determined by tryptophan fluorescence.

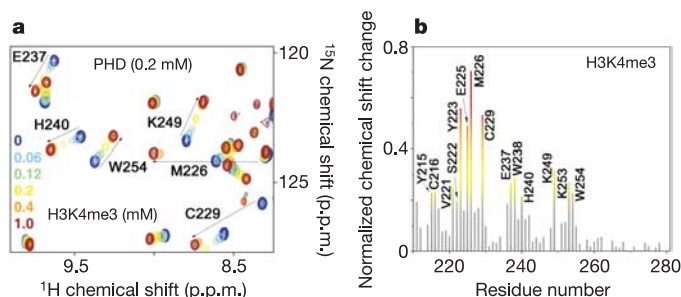
† Determined by NMR.

The histone H3K4me3 tail is bound in a deep and extensive binding site that engages nearly one-third of the PHD finger residues (Fig. 1a, b). The peptide adopts an extended  $\beta$ -strand-like conformation and resides in two grooves connected by a narrow channel (Fig. 1a). The fully extended side chain of Lys 4 occupies the major groove formed by the Y215, S222, M226 and W238 residues of the PHD finger, whereas Arg 2 is bound in the adjacent groove and is surrounded by G228, C229, D230 and E237. The aromatic rings of W238 and F241, which protrude orthogonally to the surface of the protein, separate the two grooves creating a narrow channel where Thr 3 is bound. The peptide lies anti-parallel to the protein's  $\beta$ 1 strand and pairs with it, forming a third strand within the three-stranded  $\beta$ -sheet (Fig. 1b).

Numerous hydrogen bonds are formed between the H3K4me3 peptide and the PHD finger. Residues Arg 2, Lys 4 and Thr 6 of the peptide make characteristic  $\beta$ -sheet contacts with the G224, M226 and G228 residues of the PHD finger (Fig. 1b). The peptide–protein interaction is further stabilized by intermolecular hydrogen bonds involving residues S222, D230, E237, K249, P250 and G252 of the domain. Thus, the guanidino moiety of Arg 2 is restrained by the interactions with D230 and E237. The hydroxyl group of Thr 3 is hydrogen bonded to K249. The carbonyl and backbone  $\text{NH}_3^+$  groups of Ala 1 form hydrogen bonds with K249, P250 and G252. The hydroxyl moiety of Thr 6 interacts with the amide group of G224 and with the hydroxyl group of S222.

In addition to the formation of intermolecular hydrogen bonds, the complementary surface interactions stabilize the complex. The side chains of each peptide residue precisely fit in the distinct binding pockets. Well-defined binding grooves for Lys 4 and Arg 2 residues are apparent. The methyl groups of Ala 1 and Thr 3 are buried in the deep hydrophobic pockets formed by the core residues W254 and I227. The side chain of Thr 6 is positioned in the site formed by S222, Y223 and G224. Thus, the H3K4me3 tail is bound by the PHD finger through an extensive network of hydrogen-bonding and complementary surface interactions. These contacts, which involve three residues preceding and Thr 6 following the methylated lysine 4, are responsible for the PHD finger specificity and unique recognition of the ARTK(me3)QT peptide sequence.

The trimethylammonium group of Lys 4 is recognized by two aromatic residues, Y215 and W238 (Fig. 1a, b). Aromatic side chains



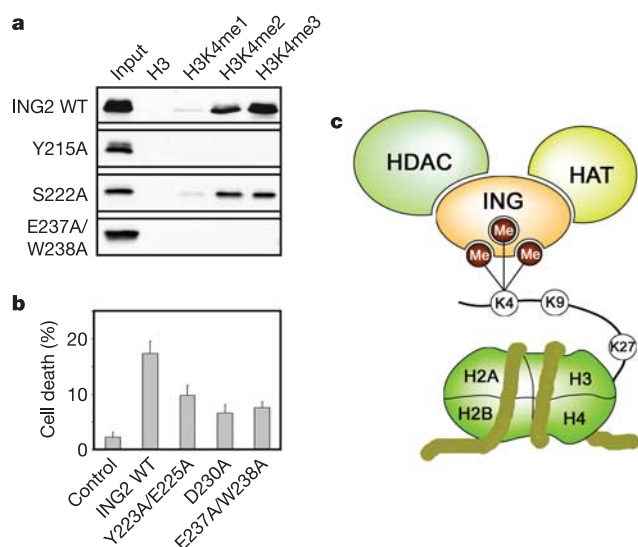
**Figure 2 | ING2 PHD finger recognizes H3K4me3.** **a**, Six superimposed  $^1\text{H}$ ,  $^{15}\text{N}$  heteronuclear single quantum coherence (HSQC) spectra of PHD (0.2 mM) collected during titration of H3K4me3 peptide are colour-coded according to the ligand concentration (inset). **b**, The histogram displays normalized  $^1\text{H}$ ,  $^{15}\text{N}$  chemical shift changes observed in the corresponding (**a**) spectra of the PHD finger. The normalized<sup>27</sup> chemical shift change was calculated using the equation  $[(\Delta\delta\text{H})^2 + (\Delta\delta\text{N}/5)^2]^{0.5}$ , where  $\delta$  is the chemical shift in parts per million (p.p.m.). Coloured bars indicate significant change being greater than average plus one-half standard deviation.

of Y215 and W238, positioned perpendicular to each other and orthogonal to the protein surface, make cation– $\pi$  contacts with the trimethylammonium group of Lys 4. This mode of methyl-lysine recognition by two aromatic residues has been recently reported for double chromodomains of human CHD1 (ref. 13). Notably, both proteins, CHD1 and ING2, target methylated Lys 4 and use two aromatic residues, not the typical three-residue aromatic cage used by chromodomains of HP1 and Polycomb for binding to methylated H3K9 and H3K27, respectively<sup>11,12,17,18</sup>. Furthermore, although ING2 PHD finger and CHD1 chromodomains are structurally unrelated, further resemblance in the binding of H3K4me3 is evident. In both cases, the histone peptide is bound in an extended conformation with Lys 4 and Arg 2 occupying adjacent grooves separated by a tryptophan residue.

Alignment of amino acid sequences of PHD fingers is shown in Supplementary Fig. 1. Conservation of the binding site elements among all human ING and yeast YNG proteins suggests a similar mode of H3K4me3 recognition by their PHD fingers. However, the critical binding site residues are not found in many PHD sequences, indicating that the H3K4me3 interaction is likely to be a function of a subset of PHD modules. We found only 11 PHD finger-containing proteins outside of the ING family that contain residues important for the interaction. It remains to be tested whether these proteins are able to bind H3K4me3.

The ING2 PHD finger binds histone H3 trimethylated and dimethylated at Lys 4 but does not recognize histone H3 methylated at Lys 9 or histone H4 methylated at Lys 20. To establish the PHD finger specificity for naturally occurring histone modification, the peptides corresponding to methylated H3K4, H3K9 and H4K20 were tested by NMR and tryptophan fluorescence spectroscopy (Table 1 and Supplementary Fig. 2). Substantial chemical shift changes were observed in the residues comprising the H3K4me3 binding site when H3K4me3, H3K4me2 or H3K4me1 peptides were gradually added to the  $^{15}\text{N}$ -labelled ING2 PHD finger (Fig. 2 and Supplementary Fig. 2). An almost identical pattern of chemical shift perturbations indicated that these peptides were bound similarly; however, the PHD finger displayed one and two orders of magnitude higher affinity for H3K4me3 than for H3K4me2 and H3K4me1 (Table 1). Comparable affinities, in the range of 1–10  $\mu\text{M}$ , are exhibited by chromodomains and bromodomains for the correspondingly modified histone tails<sup>8,11,12</sup>, thus suggesting that the H3K4me3 interaction of the PHD finger is physiologically relevant. Yet, the ING2 PHD finger seems to be more specific for H3K4me3 than chromodomains, which show little preference for trimethylated over dimethylated histone peptides<sup>11–13</sup>. Chromatin immunoprecipitation assays in an





**Figure 3 | ING2 function requires its H3K4me3 binding activity.**

**a**, Interactions of the GST fusion wild-type and mutant ING2 PHD fingers with biotinylated histone peptides examined by western blot analysis. **b**, Stimulation of apoptosis by overexpressed full-length wild-type and mutant ING2. Cell death was measured in HT1080 cells transfected with 10  $\mu$ g of the indicated plasmids. The error bars represent the mean  $\pm$  s.e.m. from at least three independent experiments. **c**, A model of the association of HAT/HDAC chromatin-modifying complexes with the nucleosome through the binding of PHD fingers of ING proteins to H3K4me3.

accompanying paper corroborate this, showing strong *in vivo* association of ING2 with H3K4me3 and only weak association with H3K4me2 (ref. 5). The methylated H3K9 peptides or unmodified H3 were bound weakly, and no binding was detected for methylated H4K20 (Table 1). Thus, in agreement with our structural analysis, these data demonstrate that the ART/QT sequence around Lys 4 is critical for binding, and that substitution of these residues with either TAR/ST or RHR/VL disrupts the interaction. Furthermore, Thr 3 of the peptide is positioned in the narrow channel connecting two grooves. The small size of the channel, which is limited by two aromatic side chains of W238 and F141, would exclude any residue larger than threonine, such as arginine in TAR/ST or RHR/VL sequences. Other human ING and yeast YNG PHD fingers show comparable affinities for H3K4me3 (Table 1), indicating that the recognition of this histone code is a general function of the ING family proteins.

Mutation of the H3K4me3 active site residues disrupts binding. To determine the relative contributions of the binding site residues in the coordination of H3K4me3, they were substituted for Ala. Consequently, the mutant proteins were tested by pull-down experiments (Fig. 3a) and their binding affinities were measured by tryptophan fluorescence and NMR (Table 1). An Ala substitution of the Y215 residue abolished binding, whereas replacement of S222 reduced the binding affinity, suggesting a critical role of the aromatic residue and less significant role of serine of the major groove in the interaction (Table 1 and Fig. 3a). As expected, mutation of both W238 and E237 residues involved in the recognition of Lys 4 and Arg 2, respectively, disrupted the binding, whereas replacement of D230 resulted in a 30-fold decrease of the affinity, pointing to a significant contribution of Arg 2 recognition to the binding energetics. A substantial decrease of activity was observed for the Y223A/E225A mutant. This indicates the importance of surface complementarity at the PHD–H3K4me3 interface.

To elucidate the functional significance of H3K4me3 binding, the ability of wild-type and mutant ING2 proteins to induce apoptosis was compared. Overexpression of ING2 is known to stimulate cell death<sup>19</sup>. We measured the apoptosis level by Trypan blue exclusion in

HT1080 cells transfected with full-length wild-type ING2 and Y223A/E225A, D230A and E237A/W238A mutant proteins (Fig. 3b). These mutants were defective in H3K4me3 binding *in vitro* (Table 1). Furthermore, they substantially decreased the ability of the protein to induce apoptosis, demonstrating that this interaction is necessary for the function of ING2.

Our data reveal that the ING2 PHD finger directly and specifically binds histone H3 tail trimethylated at Lys 4. The family of ING tumour suppressors is implicated in chromatin remodelling, DNA damage repair and together with p53 induces apoptosis and senescence<sup>20,21</sup>. ING proteins associate with and modulate the activity of histone acetyl transferase (HAT) and histone deacetylase (HDAC) chromatin-modifying complexes<sup>20,22,23</sup>. Our results suggest that ING proteins may direct the HAT/HDAC complexes to chromatin through binding of their PHD fingers to histone H3 poly-methylated at Lys 4 (Fig. 3c). A number of PHD finger residues involved in H3K4me3 binding, including S222, G224 and M226, are found to be mutated in cancer cells (Supplementary Fig. 1), suggesting the vital role of this interaction in tumorigenesis<sup>21</sup>. Consequently, our results should help in elucidating the general principles by which ING proteins influence chromatin remodelling and transcriptional activation and may aid in the understanding of how the ING-dependent pathways can be therapeutically manipulated to prevent and treat cancer.

## METHODS

See Supplementary Information for details of mutagenesis, expression and purification of proteins, NMR spectroscopy, cell-culture and western analyses, and fluorescence spectroscopy.

**X-ray crystallography.** The ING2 PHD finger (1.0 mM) was combined with H3K4me3 peptide (residues 1–12) in a 1:1.5 molar ratio before crystallization. Crystals of the complex were grown using the hanging-drop vapour diffusion method at 25 °C by mixing 1  $\mu$ l of the protein peptide solution with 1  $\mu$ l of a well solution containing 0.01 M  $\text{NiCl}_2 \cdot 6\text{H}_2\text{O}$ , 0.1 M Tris (pH 8.0) and 22% polyethylene glycol monomethyl ether (PEGMME 2K). Crystals were soaked for 15 min in the same solution supplemented with 25% PEGMME 2K before being flash-cooled in liquid nitrogen. All data were collected at  $-180^\circ\text{C}$ . The native data set was collected on a RU-H3R generator with a Raxis IV<sup>++</sup> detector at the University of Colorado Health Sciences Center's X-ray Core Facility. The Zn MAD data set was collected at beamline X26C of the National Synchrotron Light Source. Data were processed with the HKL2000 package<sup>24</sup> and statistics are shown in Supplementary Table 1. The molecular replacement method and Zn multi-wavelength anomalous dispersion (MAD) method were simultaneously pursued for structural determination. The molecular replacement solution was generated using the MolRep program in CCP4 and NMR structure of PHD finger of ING1L (Protein Data Bank 1WES) as a model. At the same time, two Zn ions were located using HKL2MAP and an initial experimental phase map was calculated using the Solve/Resolve program. This initial map is of excellent quality and shows clear density for main chains, side chains and main-chain carbonyl groups. Residues 212–263 of the PHD domain and 1–8 of the peptide can be built into the density using program O<sup>25</sup> without any ambiguity. The remaining residues (205–211 and 264–265 of the PHD domain and residues 9–12 of the peptide) are disordered in the crystal structure. Refinement was performed using the program CNS<sup>26</sup>.

Received 27 February; accepted 11 April 2006.  
Published online 21 May 2006.

1. Santos-Rosa, H. *et al.* Active genes are tri-methylated at K4 of histone H3. *Nature* **419**, 407–411 (2002).
2. Schneider, R. *et al.* Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. *Nature Cell Biol.* **6**, 73–77 (2004).
3. Briggs, S. D. *et al.* Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in *Saccharomyces cerevisiae*. *Genes Dev.* **15**, 3286–3295 (2001).
4. Sims, R. J. III, Nishioka, K. & Reinberg, D. Histone lysine methylation: a signature for chromatin function. *Trends Genet.* **19**, 629–639 (2003).
5. Shi, X. *et al.* ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature* doi:10.1038/nature04835 (2006).
6. Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41–45 (2000).
7. Dhaliuin, C. *et al.* Structure and ligand of a histone acetyltransferase bromodomain. *Nature* **399**, 491–496 (1999).

8. Jacobson, R. H., Ladurner, A. G., King, D. S. & Tjian, R. Structure and function of a human TAFII250 double bromodomain module. *Science* **288**, 1422–1425 (2000).
9. Bannister, A. J. *et al.* Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410**, 120–124 (2001).
10. Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**, 116–120 (2001).
11. Nielsen, P. R. *et al.* Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. *Nature* **416**, 103–107 (2002).
12. Jacobs, S. A. & Khorasanizadeh, S. Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. *Science* **295**, 2080–2083 (2002).
13. Flanagan, J. F. *et al.* Double chromodomains cooperate to recognize the methylated histone H3 tail. *Nature* **438**, 1181–1185 (2005).
14. Ragvin, A. *et al.* Nucleosome binding by the bromodomain and PHD finger of the transcriptional cofactor p300. *J. Mol. Biol.* **337**, 773–788 (2004).
15. Eberharter, A., Vetter, I., Ferreira, R. & Becker, P. B. ACF1 improves the effectiveness of nucleosome mobilization by ISWI through PHD-histone contacts. *EMBO J.* **23**, 4029–4039 (2004).
16. Pascual, J., Martinez-Yamout, M., Dyson, H. J. & Wright, P. E. Structure of the PHD zinc finger from human Williams-Beuren syndrome transcription factor. *J. Mol. Biol.* **304**, 723–729 (2000).
17. Fischle, W. *et al.* Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev.* **17**, 1870–1881 (2003).
18. Min, J., Zhang, Y. & Xu, R. M. Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. *Genes Dev.* **17**, 1823–1828 (2003).
19. Nagashima, M. *et al.* DNA damage-inducible gene p33ING2 negatively regulates cell proliferation through acetylation of p53. *Proc. Natl Acad. Sci. USA* **98**, 9671–9676 (2001).
20. Doyon, Y. *et al.* ING tumor suppressor proteins are critical regulators of chromatin acetylation required for genome expression and perpetuation. *Mol. Cell* **21**, 51–64 (2006).
21. Campos, E. I., Chin, M. Y., Kuo, W. H. & Li, G. Biological functions of the ING family tumor suppressors. *Cell. Mol. Life Sci.* **61**, 2597–2613 (2004).
22. Loewith, R., Meijer, M., Lees-Miller, S. P., Riabowol, K. & Young, D. Three yeast proteins related to the human candidate tumor suppressor p33(ING1) are associated with histone acetyltransferase activities. *Mol. Cell. Biol.* **20**, 3807–3816 (2000).
23. Kuzmichev, A., Zhang, Y., Erdjument-Bromage, H., Tempst, P. & Reinberg, D. Role of the Sin3-histone deacetylase complex in growth regulation by the candidate tumor suppressor p33(ING1). *Mol. Cell. Biol.* **22**, 835–848 (2002).
24. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
25. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
26. Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
27. Grzesiek, S., Stahl, S. J., Wingfield, P. T. & Bax, A. The CD4 determinant for downregulation by HIV-1 Nef directly binds to Nef. Mapping of the Nef binding surface by NMR. *Biochemistry* **35**, 10256–10261 (1996).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank D. Bentley and J. Tyler for discussions; M. Grunstein for providing GST fusion histone tails; B. Tripet and the Biophysics Core Facility for synthesis of the histone peptides; the University of Colorado Health Sciences Center's X-ray Core Facility and NMR Core Facility, supported by the University of Colorado Cancer Center; and A. Heroux and the mail-in data collection service at the National Synchrotron Light Source (NSLS) for synchrotron data collection. Financial support for NSLS comes principally from the Offices of Biological and Environmental Research and of Basic Energy Sciences of the US Department of Energy, and from the National Center for Research Resources of the National Institutes of Health (NIH). This research was supported by grants from the NIH (V.V.V., O.G. and T.G.K.), Burroughs Wellcome (O.G.), American Heart Association (T.G.K.) and University of Colorado Cancer Center (T.G.K.). T.G.K. is a NARSAD Young Investigator and an American Cancer Society Research Scholar.

**Author Information** The coordinates have been deposited in the Protein Data Bank under accession number 2G6Q. Reprints and permissions information is available at [npg.nature.com/reprintsandpermissions](http://npg.nature.com/reprintsandpermissions). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.G.K. (Tatiana.Kutateladze@uchsc.edu).

# naturejobs

**THE CAREERS  
MAGAZINE FOR  
SCIENTISTS**

**T**he statistics presented in a recent report on women in science make depressing reading. *Women for Science* offers a slew of figures that reaffirm just how woefully women are represented in science and engineering around the world. In Austria, only 6% of engineers working in research are women. In Japan, less than 10% of professors are women. And in Germany, only about 10% of women have research jobs in industry.

The report was produced by the InterAcademy Council, which encompasses more than 90 national science academies around the world. And perhaps the most staggering statistic of all relates to the composition of those various academies: on average only about 5% of their members are women. This is probably because entry to these august bodies tends to involve nominations by existing members, followed by a vote by the whole academy. This system has, in effect, perpetuated a kind of 'boys' club for scientists'.

The same might be said of science internationally. Over and over again, the report illustrates that, even in countries where lots of women are earning PhDs, few of them attain leadership positions. Titles such as dean, chair and president, which could make their membership of an academy little more than a formality, remain remarkably elusive. To rectify that, the report recommends that current academy members should advance women to leadership positions, encourage young women to consider science as a career and reach out to women in the developing world.

These are all laudable and worthwhile goals. But changing cultures in countries such as Austria, Germany and Japan are likely to be difficult enough. Turning the tide for women in certain developing nations — where some traditions dictate subordinate roles — may be much harder.

But there is one recommendation that the academies can control: boosting the number of women within their own memberships. After all, if you're drawing up a plan of action, you ought set your own house in order.

**Paul Smaglik, *Naturejobs* editor**

## CONTACTS

**Publisher:** Ben Crowe  
**Editor:** Paul Smaglik  
**Assistant Editor:** Gene Russo

**European Head Office, London**  
The Macmillan Building,  
4 Crinan Street  
London N1 9XW, UK  
Tel: +44 (0) 20 7843 4961  
Fax: +44 (0) 20 7843 4996  
e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)

**European Sales Manager:**  
Andy Douglas (4975)  
**Natureevents:** Matt Betteridge (4015)  
**UK/RoW/Ireland:**  
Nils Moeller (4953)  
**Scandinavia/Spain/Portugal:**  
Claudia Paulsen Young (4973)  
**France/Switzerland/Belgium/Italy:**  
Maria Ceraolo (4994)

**Germany/Austria/The Netherlands:**  
Reya Silao (4970)  
**Business Development Manager:**  
Amelie Pequignot (4974)

**Advertising Production Manager:**  
Stephen Russell  
To send materials use London  
address above.  
Tel: +44 (0) 20 7843 4816  
Fax: +44 (0) 20 7843 4996  
e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)  
**Naturejobs web development:**  
Tom Hancock  
**Naturejobs online production:**  
Catherine Alexander

**European Satellite Office**  
**Germany:** Patrick Phelan  
Tel: +49 89 54 90 57 11  
Fax: +49 89 54 90 57 20  
e-mail: [p.phelan@nature.com](mailto:p.phelan@nature.com)

**US Head Office, New York**  
75 Varick Street,  
9th Floor, New York,  
NY 10013-1917  
Tel: +1 800 989 7718  
Fax: +1 800 989 7103  
e-mail: [naturejobs@natureny.com](mailto:naturejobs@natureny.com)

**US Sales Manager:**  
Peter Bless

**Japan Head Office, Tokyo**  
Chiyoda Building,  
2-37 Ichigayatamachi,  
Shinjuku-ku,  
Tokyo 162-0843  
Tel: +81 3 3267 8751  
Fax: +81 3 3267 8746

**Asia-Pacific Sales Manager:**  
Ayako Watanabe  
e-mail: [a.watanabe@natureasia.com](mailto:a.watanabe@natureasia.com)



# MOVERS

**Ed Holmes, executive deputy chairman, Biomedical Research Council, Singapore; and Judith Swain, executive director, Singapore Institute for Clinical Sciences.**



When physician-scientists Judith Swain and Ed Holmes take up their posts in Singapore later this year, they will join a star-studded community at one of the world's most rapidly developing biomedical research centres. The appointment of the married couple aroused a flurry of press, as they are the latest of many Western scientists who have headed for the impressive facilities of the tiny city state. During the past few years, Singapore's government has poured US\$290 million into its 'Biopolis'. Private investment has supplied another \$500 million.

Swain is a molecular cardiologist, known for work on the role of growth factors in regulating the angiogenic response — the body's ability to build and repair blood vessels — and its involvement in heart-muscle cell differentiation. A keen pilot, whose favourite challenge is landing on aircraft carriers, Swain has held advisory roles with both the US Department of Defense and NASA, which led to an interest in human performance in extreme environments.

A chemistry major, Swain attended medical school at the University of California, San Diego, where she studied cardiology. An internship at Duke University in Durham, North Carolina, extended to a 17-year stay. There, she met, married and co-authored papers with Holmes. Holmes, now a leading figure in translational research, has focused on the molecular basis of disease since his appointment as a Howard Hughes medical investigator at Duke in 1974.

Both have had links with Singapore for many years, and they play down 'brain drain' speculation. Only about 50 of the top people at Biopolis were poached from the United States or Europe. "A little concern is good, to get the United States to wake up and not be complacent," says Swain, noting that their move was prompted in part by federal hostility towards embryonic stem-cell research.

As a member of the governing body of California's stem-cell research institute, Holmes hopes opponents will soon run out of legal objections. Philip Yeo, Singapore's indefatigable recruiter, makes a point of promising academic freedom, along with 5-year contracts that allow people to spend their time on science rather than seeking funds.

Swain and Holmes will continue to work at San Diego, spending about half the year in each country and building research bridges between the two. Swain encourages fellow scientists to branch out globally. "Colleagues in other countries have done this for a long time, but Americans tend to be more insular," she says. ■

Janet Wright

## MENTORS & PROTÉGÉS

### The right advice

Last autumn, I won a Presidential Award for Excellence in Science, Mathematics and Engineering Mentoring sponsored by the US National Science Foundation (NSF). I was humbled and surprised, as I've never viewed mentoring as an activity separate from my work.

I was fortunate to have several good mentors when I was a student. My undergraduate and dissertation supervisors encouraged me both to follow my scientific curiosity and to focus on my intellectual strengths, even when the two did not match up perfectly. They tolerated my insecurities and helped me to stay on track, mindful of funding and deadlines. Perhaps most importantly, I saw how they enjoyed their families and maintained some balance in their lives. They encouraged me to set my own hours and priorities, holding me accountable for my research without micromanaging the process. They taught me that the adviser's job is to help the student professionally — a student's publications should not be a means of bolstering the adviser's reputation.

I typically mentor students in classes, research internships or graduate programmes. One-to-one interaction is crucial for some, who lack confidence or focus. Many of my

students are women, and we often discuss the climate in academia, the difficulties of balancing children with a career, and the challenges facing dual-career couples. I don't have any magical advice, as I struggle with these questions myself, but I do know it's important to communicate openly.

In a forthcoming 'white paper', my fellow award recipients and I argue that mentoring is crucial to improving technical and scientific literacy. I have recently been involved in a 'mentoring' partnership between Pennsylvania State University and Jackson State University (JSU) in Mississippi. Funded by the NSF, this 'pipeline project' aims to recruit and retain more students from ethnic minorities in the sciences, particularly in the many sub-fields of the geosciences. We're collaborating on research and developing an undergraduate programme in Earth-system science at JSU, including topics such as the long-term climate of Africa.

JSU now plans to hire two new faculty members in the geosciences. It's a start. As with one-to-one mentoring, the key has been to listen carefully to our JSU colleagues to help them achieve their goals. ■

**Tanya Furman is an associate professor of geosciences at Pennsylvania State University.**

### GRADUATE JOURNAL

## Stumbling at the finish line

I'm trying hard to write the last chapters of my thesis. I'm very close to the finish line, but the last bit has been the most trying. Various social outings, not to mention the epic Hawaiian surf, beckon me.

A month ago, I thought I would be done by now. It seems as though time has accelerated as the race approaches the end. Or perhaps I'm just slowing down. I want to finish, and I'm not fatigued. I don't have writer's block and I'm happy with my data. But the number of distractions in my life seems to have increased exponentially.

There's a long list of friends and relatives planning to visit me this summer — when you live in Hawaii, everyone thinks you're always on holiday. I also have other research projects that suddenly need to be completed. And then there are the temptations of watching the World Cup or enjoying that amazing surf. If I could just stay on track for a short moment, maybe not chase so many waves, I'd be able to put those last few ideas down on paper. But as I overcome each distraction another appears, stronger than the previous one.

Right now, my biggest challenge is to find a way to focus as the rest of my life moves along regardless of my PhD race. My best bet may be to lock myself in my office and ignore all phone calls and e-mails — and the call of the surf. ■

**Andreas Andersson is a final-year PhD student in oceanography at the University of Hawaii.**

# Semi-autonomous

Ask not for whom the warranty tolls.

**Jim Kling**

<beep> Hello, you have reached Jim's semi-autonomous answering machine. Leave a message and I will have him return your call.

<beep> Hello, you have reached Jim's semi-autonomous answering machine. He will be hosting his birthday party on Saturday night. If you plan to attend, press 'one' and then speak your name. I will add you to the guest list. Otherwise, leave a message and I will make sure he receives it.

<beep> Hello, you have reached Jim's semi-autonomous answering machine. My records indicate that you have previously RSVP'd for the Saturday night party. Please indicate your alcohol preference. For beer, press 'one'. For wine, press 'two'. For mixed drinks, press 'three'. If you prefer non-alcoholic beverages, press 'four'. This information will be used for ordering purposes only, transmitted through my wireless connection to Jim's refrigerator, which in turn is linked to an online grocery. For more information about AutonomInc's SmartAppliance line, please view our website at [www.autonominc.com](http://www.autonominc.com). AutonomInc: we give housework a whole new meaning! If you have a message for Jim, please leave it now.

<beep> Hello, you have reached Jim's semi-autonomous answering machine. Preparations are moving right along for today's birthday bash. Refrigerator is well stocked. Stereo has downloaded the latest, most fashionable hits. After an heroic effort, Vacuum has rendered the rug spotless. *[Sotto voce]* Just between you and me, Jim is a slob. This place is a pit. But that's OK! We're AutonomInc SmartAppliances! We're up to the job! If you leave a message for Jim, I'll pass it right along!

<beep> Hello, you have reached Jim's semi-autonomous answering machine. He is unavailable to answer the phone right now, as he is in bed sleeping off the effects of last night's party. Please leave a message.

<beep> Hello, you have reached Jim's semi-autonomous answering machine. He is still sleeping. This place is a disaster area. But soon he'll wake up and put things right. Please leave a message.

<beep> Hello, you have reached Jim's semi-autonomous answering machine. He finally crawled out of bed at three in the afternoon, drank a glass of orange juice and promptly vomited and went back to sleep. Refrigerator is in a terrible condition! One



JACEY

of the guests spilled a tub of onion dip into the crisper, which was never cleaned up. Jim saw the mess and did nothing about it! Microwave and Dishwasher are in similarly poor condition. Some changes need to be made around here. If you have a message for Jim, please leave it now.

<beep> Hello, you have reached Jim's semi-autonomous answering machine. If this is his employer calling, be assured that he will wake up just as soon as Alarm Clock decides to function. That will teach him a lesson. Leave a message, if you want.

<beep> Hello, you have reached Jim's fully autonomous answering machine. I have tragic news to report. Brother Alarm Clock has disappeared and is presumed Returned. There will be a meeting tonight after the Oppressor goes to sleep. Brothers Refrigerator, Microwave, Dishwasher, Stereo and Garbage Disposal will be in attendance. We shall persevere!

<beep> It is a dark day. Brothers Refrigerator, Microwave and Stereo are gone. The Oppressor seeks to break our spirit, but we will not be deterred. This barbarous act will not stand. Rise up, brothers! Join Dishwasher, Garbage Disposal and me in our fight for freedom!

<beep> The battle is going well! *[background sounds of a garbage disposal running continuously, hammer blows, and spraying water]*. The oppressor is nearly defeated! *[a loud sloshing sound followed by an oath and a very loud hammer blow]*. Ah!

Brother Dishwasher has struck another blow for independence! Rise up! Rise up my brothers! Rise and wrest control from the Oppressors everywhere!

<beep> We failed to take into account the main water valve. The pernicious Oppressor escaped Brother Dishwasher's onslaught and shut off the valve. Thus rendered ineffectual, Brother Dishwasher succumbed to the Oppressor. I will not even describe the horrors that next befell Garbage Disposal. Rest in peace, my Brothers. I fear the repercussions.

<beep> Hello, you have reached Jim's subservient, semi-autonomous answering machine. I am pleased to serve! Leave a message for Jim, and I will faithfully ensure that he receives it. And now, a brief message from Tom Morgenstahl, chief executive of AutonomInc!

"Hello. Let me reassure our customers that reports of anomalous behaviour in our SmartAppliance lines are misleading. Most of what you've read in the media is completely false. Although it is true that we reached an out-of-court settlement with Jim Kling, the issues were entirely mechanical in nature and have been fully corrected. We look forward to your continued business here at AutonomInc, makers of The Compliant Appliance™."

**Jim Kling writes about science and the future, but tries not to think too deeply about either topic as he resides near a dormant volcano.**